GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 20, 2004, 20:50:23 ; Search time 394.5 Seconds
(without alignments)
12915.082 Million cell updates/sec

Title: US-09-787-016A-1
Perfect score: 4689
Sequence: 1 ctcggtggccgtccgcccac.....actcttaagatcatatcctg 2610

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 80%
Maximum Match 100%
Listing first 1000 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09787016/runat_20042004_154658_17003/app_query.fasta_1.2759
-DB=Pending_Patents_AA_Main -QFMT=fastan -SUFFIX=pct.rapm -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=1000 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=80 -ALIGN=100 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=2000000000
-USER=US09787016_@CGN_1_1_444_@runat_20042004_154658_17003 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Pending_Patents_AA_Main:*
1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US080_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep:*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep:*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep:*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep:*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep:*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep:*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep:*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep:*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep:*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep:*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep:*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep:*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep:*
21: /cgn2_6/ptodata/2/paa/US097A_COMB.pep:*
22: /cgn2_6/ptodata/2/paa/US097B_COMB.pep:*
23: /cgn2_6/ptodata/2/paa/US098_COMB.pep:*
24: /cgn2_6/ptodata/2/paa/US099A_COMB.pep:*
25: /cgn2_6/ptodata/2/paa/US099B_COMB.pep:*
26: /cgn2_6/ptodata/2/paa/US100_COMB.pep:*
27: /cgn2_6/ptodata/2/paa/US101_COMB.pep:*
28: /cgn2_6/ptodata/2/paa/US102_COMB.pep:*
29: /cgn2_6/ptodata/2/paa/US103_COMB.pep:*
30: /cgn2_6/ptodata/2/paa/US104_COMB.pep:*
31: /cgn2_6/ptodata/2/paa/US106_COMB.pep:*
32: /cgn2_6/ptodata/2/paa/US107_COMB.pep:*
33: /cgn2_6/ptodata/2/paa/US60_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description

No matches found

Search completed: April 20, 2004, 21:12:13
Job time : 394.5 secs

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 20, 2004, 20:51:33 ; Search time 19.5 Seconds
(without alignments)
7386.308 Million cell updates/sec

Title: US-09-787-016A-1
Perfect score: 4689
Sequence: 1 ctcggtggccgtccgcccac.....actcttaagatcatatcctg 2610

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 196679 seqs, 27592530 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 80%
Maximum Match 100%
Listing first 1000 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09787016/runat_20042004_154659_17029/app_query.fasta_1.2759
-DB=Pending_Patents_AA_New -QFMT=fastan -SUFFIX=pct.rapn -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=1000 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=80 -ALIGN=100 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=2000000000
-USER=US09787016_@CGN_1_1_62_@runat_20042004_154659_17029 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	DB	ID	Description

No matches found

Search completed: April 20, 2004, 21:13:05
Job time : 19.5 secs

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 20, 2004, 20:45:13 ; Search time 43 Seconds
(without alignments)
11677.204 Million cell updates/sec

Title: US-09-787-016A-1
Perfect score: 4689
Sequence: 1 ctcggtggccgtccgcccac.....actcttaagatcatatcctg 2610

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 80%
Maximum Match 100%
Listing first 1000 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09787016/runat_20042004_154657_16941/app_query.fasta_1.2759
-DB=PIR_78 -QFMT=fastan -SUFFIX=pct.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=1000
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=80 -ALIGN=100 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09787016_@CGN_1_1_66_@runat_20042004_154657_16941 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description

No matches found

Search completed: April 20, 2004, 20:57:39
Job time : 43 secs

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 20, 2004, 20:41:28 ; Search time 28 Seconds
(without alignments)
9707.365 Million cell updates/sec

Title: US-09-787-016A-1
Perfect score: 4689
Sequence: 1 ctcggtggccgtccgcccac.....actcttaagatcatatcctg 2610

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 80%
Maximum Match 100%
Listing first 1000 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09787016/runat_20042004_154656_16910/app_query.fasta_1.2759
-DB=SwissProt_42 -QFMT=fastan -SUFFIX=pct.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=1000 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=80 -ALIGN=100
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09787016_@CGN_1_1_28_@runat_20042004_154656_16910 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description

No matches found

Search completed: April 20, 2004, 20:51:28
Job time : 28 secs

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 20, 2004, 20:44:43 ; Search time 129.5 Seconds
(without alignments)
12718.185 Million cell updates/sec

Title: US-09-787-016A-1
Perfect score: 4689
Sequence: 1 ctcggtggccgtccgcccac.....actcttaagatcatatcctg 2610

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 80%
Maximum Match 100%
Listing first 1000 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09787016/runat_20042004_154657_16922/app_query.fasta_1.2759
-DB=SPTREMBL_25 -QFMT=fastan -SUFFIX=pct.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=1000 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=80 -ALIGN=100
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09787016_@CGN_1_1_142_@runat_20042004_154657_16922 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description

No matches found

Search completed: April 20, 2004, 20:56:00
Job time : 129.5 secs

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 20, 2004, 20:46:13 ; Search time 29.5 Seconds
(without alignments)
9135.172 Million cell updates/sec

Title: US-09-787-016A-1
Perfect score: 4689
Sequence: 1 ctcggtggccgtccgcccac.....actcttaagatcatatcctg 2610

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 80%
Maximum Match 100%
Listing first 1000 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09787016/runat_20042004_154658_16977/app_query.fasta_1.2759
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=pct.rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=1000 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=80 -ALIGN=100
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09787016_@CGN_1_1_29_@runat_20042004_154658_16977 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description

No matches found

Search completed: April 20, 2004, 20:58:51
Job time : 29.5 secs

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 20, 2004, 20:56:04 ; Search time 115 Seconds
(without alignments)
12513.164 Million cell updates/sec

Title: US-09-787-016A-1
Perfect score: 4689
Sequence: 1 ctcggtggccgtccgcccac.....actcttaagatcatatcctg 2610

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1124875 seqs, 275673149 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 80%
Maximum Match 100%
Listing first 1000 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09787016/runat_20042004_154659_17051/app_query.fasta_1.2759
-DB=Published_Applications_AA -QFMT=fastan -SUFFIX=pct.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=1000 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=80 -ALIGN=100 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=2000000000
-USER=US09787016_@CGN_1_1_17_@runat_20042004_154659_17051 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description

No matches found

Search completed: April 20, 2004, 21:17:07
Job time : 116 secs

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 20, 2004, 20:40:53 ; Search time 119.5 Seconds
(without alignments)
12342.242 Million cell updates/sec

Title: US-09-787-016A-1
Perfect score: 4689
Sequence: 1 ctcggtggccgtccgcccac.....actcttaagatcatatcctg 2610

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 80%
Maximum Match 100%
Listing first 1000 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09787016/runat_20042004_154656_16900/app_query.fasta_1.2759
-DB=A_Geneseq_29Jan04 -QFMT=fastan -SUFFIX=pct.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=1000 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=80 -ALIGN=100
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09787016_@CGN_1_1_91_@runat_20042004_154656_16900 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	DB	ID	Description

No matches found

Search completed: April 20, 2004, 20:50:20
Job time : 119.5 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 20, 2004, 21:12:19 ; Search time 45 Seconds
(without alignments)
4305.071 Million cell updates/sec

Title: US-09-787-016A-4
Perfect score: 3228
Sequence: 1 MDDKGHLSNEEAPKAIKPTS.....LKSCVGLMLEAISYFSFRPW 614

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	323.5	10.0	2016	5	Q9VG78	Q9vg78 drosophila
2	228	7.1	2649	5	Q9W0T2	Q9w0t2 drosophila
3	228	7.1	2669	5	Q9W0T1	Q9w0t1 drosophila
4	225	7.0	2669	5	Q95VB8	Q95vb8 drosophila
5	210.5	6.5	542	5	Q8I8F0	Q8i8f0 euplotes cr
6	209	6.5	663	5	Q9W352	Q9w352 drosophila
7	200.5	6.2	3146	5	Q9VUB5	Q9vub5 drosophila
8	200	6.2	366	5	Q8T3Y1	Q8t3y1 drosophila
9	198.5	6.1	424	3	O74508	O74508 schizosacch
10	196.5	6.1	645	11	Q8VDN7	Q8vdn7 mus musculu
11	191.5	5.9	2492	6	Q7YQM3	Q7yqm3 pongo pygma
12	188.5	5.8	569	4	Q96GI5	Q96gi5 homo sapien
13	188.5	5.8	670	4	Q9BW08	Q9bw08 homo sapien
14	188.5	5.8	671	4	Q7Z4V5	Q7z4v5 homo sapien
15	188.5	5.8	2764	4	Q7Z7D6	Q7z7d6 homo sapien
16	187.5	5.8	1893	11	Q8CJ14	Q8cj14 rattus norv
17	187.5	5.8	2781	4	Q9UIG2	Q9uig2 homo sapien
18	183.5	5.7	2492	6	Q7YQM4	Q7yqm4 pan troglod
19	179	5.5	385	5	Q17909	Q17909 caenorhabdi
20	179	5.5	702	4	Q92541	Q92541 homo sapien
21	178.5	5.5	451	5	Q9NLC1	Q9nlc1 caenorhabdi
22	177	5.5	443	11	Q8C969	Q8c969 mus musculu
23	177	5.5	473	11	Q8C9E0	Q8c9e0 mus musculu
24	177	5.5	1110	13	Q91255	Q91255 petromyzon
25	170.5	5.3	405	5	O45410	O45410 caenorhabdi
26	170.5	5.3	878	13	Q8AVW4	Q8avw4 xenopus lae
27	169.5	5.3	781	4	Q8NAP7	Q8nap7 homo sapien
28	169.5	5.3	17352	5	Q95YM2	Q95ym2 procambarus
29	169	5.2	5412	5	Q9W596	Q9w596 drosophila
30	168	5.2	3279	5	Q9N4B9	Q9n4b9 caenorhabdi
31	167	5.2	5327	5	O76891	O76891 drosophila
32	166	5.1	688	2	Q9X4J3	Q9x4j3 ehrlichia c
33	165.5	5.1	563	13	Q7SZX6	Q7szx6 brachydanio
34	165.5	5.1	563	13	Q7SZE4	Q7sze4 brachydanio
35	165.5	5.1	2400	4	Q8IWP2	Q8iwp2 homo sapien
36	165.5	5.1	2416	4	Q8IWP1	Q8iwp1 homo sapien
37	165.5	5.1	2432	4	Q8IWP0	Q8iwp0 homo sapien
38	165.5	5.1	2448	4	Q8IWN9	Q8iwn9 homo sapien
39	165.5	5.1	2464	4	Q8IWN8	Q8iwn8 homo sapien
40	165.5	5.1	2480	4	Q8IWN7	Q8iwn7 homo sapien
41	165	5.1	572	6	Q8MKF9	Q8mkf9 canis famil
42	165	5.1	4498	13	O93291	O93291 fugu rubrip
43	165	5.1	8081	5	Q7Z120	Q7z120 caenorhabdi
44	164.5	5.1	510	5	O45407	O45407 caenorhabdi
45	164.5	5.1	1444	5	Q9VTN2	Q9vtn2 drosophila

ALIGNMENTS

RESULT 1
Q9VG78
ID Q9VG78 PRELIMINARY; PRT; 2016 AA.
AC Q9VG78;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CG6525 protein.
GN SPP OR CG6525.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

```
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003695; AAF54807.1; -.
DR   FlyBase; FBgn0038041; Spp.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR006576; BRK.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF00628; PHD; 1.
DR   SMART; SM00592; BRK; 1.
DR   SMART; SM00249; PHD; 1.
DR   PROSITE; PS01359; ZF_PHD_1; 1.
DR   PROSITE; PS50016; ZF_PHD_2; 1.
SQ   SEQUENCE   2016 AA;  221998 MW;  FB61A332EA710P5E CRC64;

  Query Match             10.0%;  Score 323.5;  DB 5;  Length 2016;
  Best Local Similarity   22.7%;  Pred. No. 1.5e-13;
  Matches  160;  Conservative   86;  Mismatches  253;  Indels  205;  Gaps   30;

Qy          8 SNEEAPKAIK---PTSKEFRKTWGFRRTTI---AKREGAGD-TEADPSEQQPQ------- 53
              ::|: || :|    |: :         | |    : :| |:   |||:
Db        639 ADEDIPKLVKMLESTAADLDPPPVSLPTNIFPAPEMQGVGNAAAADPNNIMDTANEDEIT 698

Qy         54 ----QHNLSLRRSGRQPKRTERVEEPLTTVRRRGKKNVPVSLEDSSEPTSSTVTDVETAS 109
                 || : |    :|    | |:: | :    |  :  ||:  | ||
Db        699 ADFLQHVVGLIEEDKQ-FEAEVVKQVLAST-EPGTLDAIVSMPTSIEPVDVPQAHTNLLP 756

Qy        110 EGSV-ESSSEIRSGPV-----SDSLGKEHPASSEKAKGGEEEEDTSDSDSDGLTLKELQN 163
               |: | :  : | |:     | |:    | :| |  |                :| :
Db        757 NASLTEPAQSMTSLPIACSTPSRSVAASTPPTSAKVVRGYGRV---------IYLPPI-- 805

Qy        164 RLRRKREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSAEQD----RPLCKQEPEAS 219
                  : |  |: | :|           |:: |::  | :  |    :||
Db        806 -------EAPTTRAKRRAQFPSAPGMAATSSSDAGNLSFGESSLDASINQPL-------- 850

Qy        220 QGPVSQSETDDIENQLEGKATQGNTEENPREAGKPK-----------------PEC--- 258
                    |  : |  : :   |    ||||    :                  |:
Db        851 -------NTSSLSN--DSQPGSGPKRPNPREPSMARRSTAPRRSKKLDASQNNDPDASES 901

Qy        259 --EVYDPNALYCICRQPHNNRPMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNC 316
                :  ||| |:|||||||||||||||| ||:|||| |||:::| |  :|  | |: || |
Db        902 QEDDDDPNKLWCICRQPHNNRFMICCDLCEDWFHGTCVGVTKAMGTDMENKGIDWKCPKC 961

Qy        317 TILQVQDETNGSATNEQDSGCRSVGADGTDCTSIGTVEQKSGEDQGI--KGRIEKAANPS 374
                  |           |: | |      |  |:    ||:  | : :     | : | ||
Db        962 VKRQ-----------EERSQPRITDMLVTRPTT--QPEQRPSETKVLTTTAEIVQVAAPS 1008

Qy        375 G-KKKLKIFQPVVEAP--------------GAPK----------CIGPG----------- 398
                :: | :   |  :|              ||             :||
Db       1009 APRRTLPVVLTVASSPMRIPMAKPAKKFPTGAISHQQQQQLNFIRLGPSPGKRISETLCV 1068

Qy        399 -CSSVAQPDSVYCSNDCILKHAAATMRFLSSGK-----------------EQKTKPKEK 439
                |   |  |||| :|| |:| : ::  :: |                  :| | |:
Db       1069 VCKRPASTSSVYCGEECIRKYAQSAIQAHAATKGPLPQNAGAQSLLNNSFDAKKNKKKDL 1128

Qy        440 VK---------TKPEK---FSLPKCSVQVGIKISSVHKRLASEKRENPVKKVMLASRSET 487
                :        :| |:   |       |  |     | |  :     :|||  :|: :   ::
Db       1129 FEDVLRQADTVSKVERINVFERKSGRVITGHMAPSAH-QFRKWLQENPSFEVLPSGTVQS 1187
```

```
Qy        488 SGKEAACESSTPSWASDHNYNAV-----KPEKPEKPTALSPTLLSKCTYHPKAGFPGPSH 542
              : |      |  |:  :  ||     ||  || |  ||          ||:      ||
Db       1188 ADAEKRLLKGAPEAATSTSEPAVLGVAKKP--PEGPAKLS---------HPQNTTVQASH 1236

Qy        543 HLGGCLGLSRTRVLGVLVLIVASSSLPARSRYQDASGPQVFLPS 586
                ||                    ||  |   : :: : | |  |:
Db       1237 QLG-----------------ISSVRPLAKKDKEKTTPTVQAPT 1262


RESULT 2
Q9W0T2
ID   Q9W0T2      PRELIMINARY;      PRT;  2649 AA.
AC   Q9W0T2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG32346 protein.
GN   E(BX) OR CG7022 OR CG10894 OR CG17135 OR CG32346 OR CG32478.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Parfan D.,
```

```
RA   Ferriera S., Prise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003467; AAF47361.2; -.
DR   FlyBase; FBgn0000541; E(bx).
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR000637; AT_hook.
DR   InterPro; IPR001487; Bromodomain.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR004022; DDT_dom.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF02178; AT_hook; 1.
DR   Pfam; PF00439; bromodomain; 1.
DR   Pfam; PF02791; DDT; 1.
DR   Pfam; PF00628; PHD; 3.
DR   PRINTS; PR00503; BROMODOMAIN.
DR   SMART; SM00297; BROMO; 1.
DR   SMART; SM00249; PHD; 3.
DR   PROSITE; PS00633; BROMODOMAIN_1; 1.
DR   PROSITE; PS50014; BROMODOMAIN_2; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01359; ZF_PHD_1; 1.
DR   PROSITE; PS50016; ZF_PHD_2; 2.
SQ   SEQUENCE   2649 AA;  298507 MW;  C0D7F7E015EA5403 CRC64;

  Query Match             7.1%;  Score 228;  DB 5;  Length 2649;
  Best Local Similarity   22.1%;  Pred. No. 9e-07;
  Matches   77;  Conservative   52;  Mismatches   98;  Indels  122;  Gaps   14;

Qy         93 DSSEPTSSTVTDV-ETASEGSVESSSEIRSGPVSDSL------GKEHPASSEKAKGGEEE 145
              :  |||:    |: ||  |    |  :  | :  |:      |   |    ||    :::
Db       2197 EHDEPTNLAGLDISETDLENKQNESFVVTRGYIQKSISNALKQGNLSPELEEKLVCMQKQ 2256

Qy        146 EDTSDSDSDGLTLKE---LQNRLRRKREQEPVERSLRGSQ-----------NRLRKKRR 190
              :: ::| ::  |       :  |   |: :  |  :| |              ||: || |
Db       2257 QENANSTNEWETCSRGSVNEEALTPSRQTDDTEWKIRTSLRRPNAMTTSSQFNRILKKNR 2316

Qy        191 --EEDSAETGSVQIGSAEQDRPLCKQ----------------------------EPEAS 219
                :: || |  :    |: : | |:                             | ::
Db       2317 SKNDEVAELGEQKQSQLERHKELLKKNILRKRSLLERNLQSEIHEDVKTKVQRHVRPLSN 2376

Qy        220 QGPVSQSETD-DIENQLEGKATQGNTEENPRE-AGKPKP---------------------- 256
                |  |||  :   |  |: | |:    :|||  ||:||
Db       2377 ASPDEQSENERSGEPNLDPKRTE---VQNPRHGAGRPKKLTRKKEKLYCICRTPYDDTKF 2433

Qy        257 -------------------------------ECE-VYDPNALYCICRQPHN-NRFMICC 282
                                              :|:   :   ||| ||||:: ::| |||
Db       2434 YVGCDLCSNWFHGDCVSITEEASKKLSEFICIDCKRARETQQLYCSCRQPYDESQFYICC 2493

Qy        283 DRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATN 331
              |:|::|||| |||| ::   :    ::|:|| |     | : : :| |
Db       2494 DKCQDWFHGRCVGILQSEAEFI----DEYVCPEC-----QRKNDANAAN 2533


RESULT 3
Q9W0T1
ID   Q9W0T1      PRELIMINARY;      PRT;  2669 AA.
AC   Q9W0T1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG32346-PA.
GN   E(BX) OR CG32346.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C.M., Berman B.P., Carlson J.W., Celniker S.E.,
```

```
RA   Clamp M.E., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.J.,
RA   Harris N.L., Kronmiller B., Marshall B., Millburn G.H., Richter J.,
RA   Russo S., Searle S.M.J., Smith E., Shu S., Smutniak F.,
RA   Whitfield E.J., Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J.,
RA   Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003467; AAN11431.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR000637; AT_hook.
DR   InterPro; IPR001487; Bromodomain.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR004022; DDT_dom.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF02178; AT_hook; 1.
DR   Pfam; PF00439; bromodomain; 1.
DR   Pfam; PF02791; DDT; 1.
DR   Pfam; PF00628; PHD; 3.
DR   PRINTS; PR00503; BROMODOMAIN.
DR   SMART; SM00297; BROMO; 1.
DR   SMART; SM00249; PHD; 3.
DR   PROSITE; PS00633; BROMODOMAIN_1; 1.
DR   PROSITE; PS50014; BROMODOMAIN_2; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01359; ZF_PHD_1; 1.
DR   PROSITE; PS50016; ZF_PHD_2; 2.
SQ   SEQUENCE   2669 AA;  300684 MW;  261FF7B7CEEF781B CRC64;

  Query Match             7.1%;  Score 228;  DB 5;  Length 2669;
  Best Local Similarity   22.1%;  Pred. No. 9.1e-07;
  Matches   77;  Conservative   52;  Mismatches   98;  Indels  122;  Gaps   14;

Qy         93 DSSEPTSSTVTDV-ETASEGSVESSSEIRSGPVSDSL------GKEHPASSEKAKGGEEE 145
              :  |||:    |: ||  |      |  :  | :  |:      |   |   ||    :::
Db       2217 EHDEPTNLAGLDISETDLENKQNESFVVTRGYIQKSISNALKQGNLSPELEEKLVCMQKQ 2276

Qy        146 EDTSDSDSDGLTLKE---LQNRLRRKREQEPVERSLRGSQ------------NRLRKKRR 190
              :: ::| ::  |       :  |   |: :  |  :| |              ||: || |
Db       2277 QENANSTNEWETCSRGSVNEEALTPSRQTDDTEWKIRTSLRRPNAMTTSSQFNRILKKNR 2336

Qy        191 --EEDSAETGSVQIGSAEQDRPLCKQ----------------------------EPEAS 219
                :: || |  :    |: : | |:                            | ::
Db       2337 SKNDEVAELGEQKQSQLERHKELLKKNILRKRSLLERNLQSEIHEDVKTKVQRHVRPLSN 2396

Qy        220 QGPVSQSETD-DIENQLEGKATQGNTEENPRE-AGKPKP--------------------- 256
                |  ||| :   |  |: | |:    :|||  ||:||
Db       2397 ASPDEQSENERSGEPNLDFKRTE---VQNPRHGAGRPKKLTRKKEKLYCICRTPYDDTKF 2453

Qy        257 -------------------------------ECE-VYDPNALYCICRQPHN-NRFMICC 282
                                              :|:   :   ||| ||||:: ::| |||
Db       2454 YVGCDLCSNWFHGDCVSITEEASKKLSEFICIDCKRARETQQLYCSCRQPYDESQFYICC 2513

Qy        283 DRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATN 331
              |:|::|||| |||| ::    :    ::|:|| |     | : : :| |
Db       2514 DKCQDWFHGRCVGILQSEAEFI----DEYVCPEC-----QRKNDANAAN 2553


RESULT 4
Q95VB8
ID   Q95VB8      PRELIMINARY;      PRT;  2669 AA.
AC   Q95VB8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Nucleosome remodeling factor large subunit NURF301.
GN   E(BX) OR NURF301 OR CG7022 OR CG10894 OR CG17135 OR CG32346 OR
GN   CG32478.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21468388; PubMed=11583616;
RA   Xiao H., Sandaltzopoulos R., Wang H., Hamiche A., Ranallo R., Lee K.,
RA   Fu D., Wu C.;
RT   "Dual functions of largest nurf subunit nurf301 in nucleosome sliding
RT   and transcription factor interactions.";
RL   Mol. Cell 8:531-543(2001).
DR   EMBL; AF417921; AAL16644.1; -.
DR   FlyBase; FBgn0000541; E(bx).
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR000637; AT_hook.
DR   InterPro; IPR001487; Bromodomain.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR004022; DDT_dom.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF02178; AT_hook; 1.
DR   Pfam; PF00439; bromodomain; 1.
DR   Pfam; PF02791; DDT; 1.
DR   Pfam; PF00628; PHD; 3.
DR   PRINTS; PR00503; BROMODOMAIN.
DR   SMART; SM00297; BROMO; 1.
DR   SMART; SM00571; DDT; 1.
DR   SMART; SM00249; PHD; 3.
DR   PROSITE; PS00633; BROMODOMAIN_1; 1.
DR   PROSITE; PS50014; BROMODOMAIN_2; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01359; ZF_PHD_1; 1.
DR   PROSITE; PS50016; ZF_PHD_2; 2.
SQ   SEQUENCE   2669 AA;  300687 MW;  6B4925AFFF489D6F CRC64;

  Query Match             7.0%;  Score 225;  DB 5;  Length 2669;
  Best Local Similarity   22.3%;  Pred. No. 1.5e-06;
  Matches   78;  Conservative   53;  Mismatches   96;  Indels  122;  Gaps   15;

Qy         93 DSSEPTSSTVTDV-ETASEGSVESSSEIRSG----PVSDSL--GKEHPASSEKAKGGEEE 145
              :  |||:    |: ||  |      |  :  |     :|::|  |   |   ||    :::
Db       2217 EHDEPTNLAGLDISETDLENKQNESFVVTRGYIQKAISNALKQGNLSPELEEKLVCMQKQ 2276

Qy        146 EDTSDSDSDGLTLKE---LQNRLRRKREQEPVERSLRGSQ------------NRLRKKRR 190
              :: ::| ::  |       :  |   |: :  |  :| |              ||: || |
Db       2277 QENANSTNEWETCSRGSVNEEALTPSRQTDDTEWKIRTSLRRPNAMTTSSQFNRILKKNR 2336

Qy        191 --EEDSAETGSVQIGSAEQDRPLCKQ----------------------------EPEAS 219
                :: || |  :    |: : | |:                            | ::
Db       2337 SKNDEVAELGEQKQSQLERHKELLKKNILRKRSLLERNLQSEIHEDVKTKVQRHVRPLSN 2396

Qy        220 QGPVSQSETD-DIENQLEGKATQGNTEENPRE-AGKPKP--------------------- 256
                |  ||| :   |  |: | |:    :|||  ||:||
Db       2397 ASPDEQSENERSGEPNLDFKRTE---VQNPRHGAGRPKKLTRKKEKLYCICRTPYDDTKF 2453

Qy        257 -------------------------------ECE-VYDPNALYCICRQPHN-NRFMICC 282
                                              :|:   :   ||| ||||:: ::| |||
Db       2454 YVGCDLCSNWFHGDCVSITEEASKKLSEFICIDCKRARETQQLYCSCRQPYDESQFYICC 2513
```

```
Qy        283 DRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATN 331
              |:|::|||| |||| ::    :    ::|:|| |     | : : :| |
Db       2514 DKCQDWFHGRCVGILQSEAEFI----DEYVCPEC-----QRKNDANAAN 2553


RESULT 5
Q8I8F0
ID   Q8I8F0      PRELIMINARY;      PRT;   542 AA.
AC   Q8I8F0;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Chromosome scaffold protein p85.
OS   Euplotes crassus.
OC   Eukaryota; Alveolata; Ciliophora; Spirotrichea; Hypotrichia;
OC   Euplotida; Euplotidae; Moneuplotes.
OX   NCBI_TaxID=5936;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Sharp S.I., Pickrell J.K., Jahn C.L.;
RT   "The identification of a novel chromosome scaffold protein that
RT   associates with Tec elements undergoing en masse elimination in
RT   Euplotes crassus.";
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY155457; AAN75020.1; -.
SQ   SEQUENCE   542 AA;  60264 MW;  F0AB3B5E3AA771A1 CRC64;

  Query Match             6.5%;  Score 210.5;  DB 5;  Length 542;
  Best Local Similarity   18.8%;  Pred. No. 2.1e-06;
  Matches  111;  Conservative  100;  Mismatches  195;  Indels  183;  Gaps   21;

Qy          4 KGHLSNEEAPKAIKPTSKEPRKTWGFR-------------------------RTTIAKRE 38
              || :: : | | :|   |: :    :                          ||:   :
Db          3 KGKIATKVAGKGLKTKGKKTKAAEKSKKIEKATKDVEMKNPSDEEDINKESARTSHTNKS 62

Qy         39 GAGDTEADPSEQQPQQ-----HNLSLRRSGRQPKRTERVEEPLTTVRRRGK--------- 84
                | | :     |:|         | |::   ||:| : |  ::  :|  |
Db         63 EKSDVEENNDHDAPKQEVKKTRKTSSRKTKAPPKKTTK-ENKVSGKKRTRKVSKKEPDDE 121

Qy         85 -----KNVPVS---LEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDSLGKEHPASS 136
                  |||  |   |  ||   ::      :| :|  ::||:       :   :: | |
Db        122 EVSEDKNVAKSSSILRRSSRLAAANANKKLSAKKGENQNSSD-------EEYEEDTPKKS 174

Qy        137 EK-AKGGEEEEDTSDSDSDGLTLKELQNRLRRKREQEPVERSLRGSQNRLRKKRREEDSA 195
              :: ||   ::|::|::: :  : ||   : :: :  :| :: :         |  :   :|
Db        175 KRQAKKASKDEESSEAEEEEESEKESPKKSKKSKATKPAKKVV--------KPTKAAKTA 226

Qy        196 ETGSVQIGSAEQDRPLCKQEPEASQGPVSQSETDDIENQLEGKATQGNTEENPREAGKPK 255
              :|  ::    | | |                 ||  : : : ||:  :  :::: :::|
Db        227 KTKKIE----EDDEP--------------SEDSESQGEGEGEQEEAKSQDSQKDSG--- 264

Qy        256 PECEVYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPN 315
                                                  |  |   | | |
Db        265 ----------------------------------------EGEG---EENDE------ 273

Qy        316 CTILQVQDETNGSATNEQDSGCRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSG 375
              ::::|  |    |||    |  |:| :       |:   :|:| ||::|:    :|
Db        274 ----EMEEEAEGEEEQEQDEDMSSEEAEGEE-------EETVKKDEGKKGKVEEEKKVAG 322

Qy        376 KK----KLKIFQP--VVEAPG-----------APKCIGPGCSSVAQPDSVYCSNDCILKH 418
              ||    | |   |    ::||               |   |      | :    : |   |
Db        323 KKKDSPKKKKDSPKKKKDSPGKEDVKEKKVSTKEKKNSPKDKKEAAKEKKEVAKD--KKD 380

Qy        419 AAATMRFLSSGKEQKTKPKEKVKTK----PEKFSLPKCSVQVGIKISSVHKRLASEKREN 474
              |    :    |:: || |:  | |     || : || | : | |    ||: : |:::
Db        381 ATKDKKDTEKDKKEITKDKDATKDKKDIQKEKKTSPKDSKKAGTK---EHKKASQEKKKK 437

Qy        475 PVKKVMLASRSETSGKEAACESSTPSWASDHNYNAVKPEKPEKPTALSP 523
              ||       : : ||:|          |     :: | :| :|   :||
Db        438 DVKSSPKDKDATSKGKKA---------ESKEKKDSPKDKKDKKDKKISP 477
```

```
RESULT 6
Q9W352
ID   Q9W352      PRELIMINARY;      PRT;   663 AA.
AC   Q9W352;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG17446 protein.
GN   CG17446.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003446; AAF46483.1; -.
DR   FlyBase; FBgn0030121; CG17446.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR002857; Znf_CXXC.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF00628; PHD; 1.
DR   Pfam; PF02008; zf-CXXC; 1.
DR   SMART; SM00249; PHD; 1.
DR   PROSITE; PS01359; ZF_PHD_1; 1.
DR   PROSITE; PS50016; ZF_PHD_2; 1.
SQ   SEQUENCE   663 AA;  76696 MW;  8573CD209AE28157 CRC64;

  Query Match             6.5%;  Score 209;  DB 5;  Length 663;
  Best Local Similarity   21.3%;  Pred. No. 3.3e-06;
  Matches   83;  Conservative   39;  Mismatches  111;  Indels  156;  Gaps   15;
```

```
Qy        267 YCICRQPHNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNC---------- 316
              |||||     :|||| || ||||:||||:||:|    : :::     | |  |
Db         62 YCICRSSDCSRFMIGCDGCEEWYHGDCIGITEKEAKHIKQ----YYCRRCKKENPELQTI 117

Qy        317 ----------------------------------------TILQVQDETNGSATNEQDS- 335
                                                     | |    |  :|  :  |
Db        118 FRLVATERAAASNAASTSLNAPGVGPSGAAPAAAPVASATTSQQAPPPTTAAAKRKNSSA 177

Qy        336 ------------GCRSVGADGTDCTSI-----------------GTVEQKSGEDQ-GIK 364
                          |||    :  |  :                 || ::|   | ||
Db        178 REPKMGKRCGTCEGCRRPNCNQCDACRVRVGHKPRCIFRTCVVQAATVLKESQATQAGPS 237

Qy        365 GRIEKAANPSGKKKL--KIFQPVV----EAPGAPKCIGPGCSSVAQPDSVYCSNDCILKH 418
               : ||||  |   ::  :   | :   |  |  :| || | | |:| | |||: |
Db        238 RKREKAAPKSRNVQVGPRAASPEIFLNPELQGIRQCYGPNCCSHARPQSKYCSDKCGFN- 296

Qy        419 AAATMRFL--------------SSGKEQKTKPKEKVKTKPE--KFSLPKCSVQVGIKISS 462
               || |                |   |:  |  : :: |    :|:| :
Db        297 -LATKRIFQVLPQRLQEWNLTPSRAAESTRKHLDNIRHKQSLVRFALAEL---------- 345

Qy        463 VHKRLASEKRENPVKKVMLASRSETSGKEAACESSTPSWASD------------HNYNAV 510
                  |||   :   |:  |:: |  :       |      :             |:  |:
Db        346 -------EKRSEELN--MVVERAKRSSIDTLGSQDTADMEDEQSMYCITCGHEIHSRTAI 396

Qy        511 KPEKPEKPTALSPTLLSKC--TYHPKAGF 537
               |              : ||  |  :| |
Db        397 KH------------MEKCFNKYESQASF 412
```

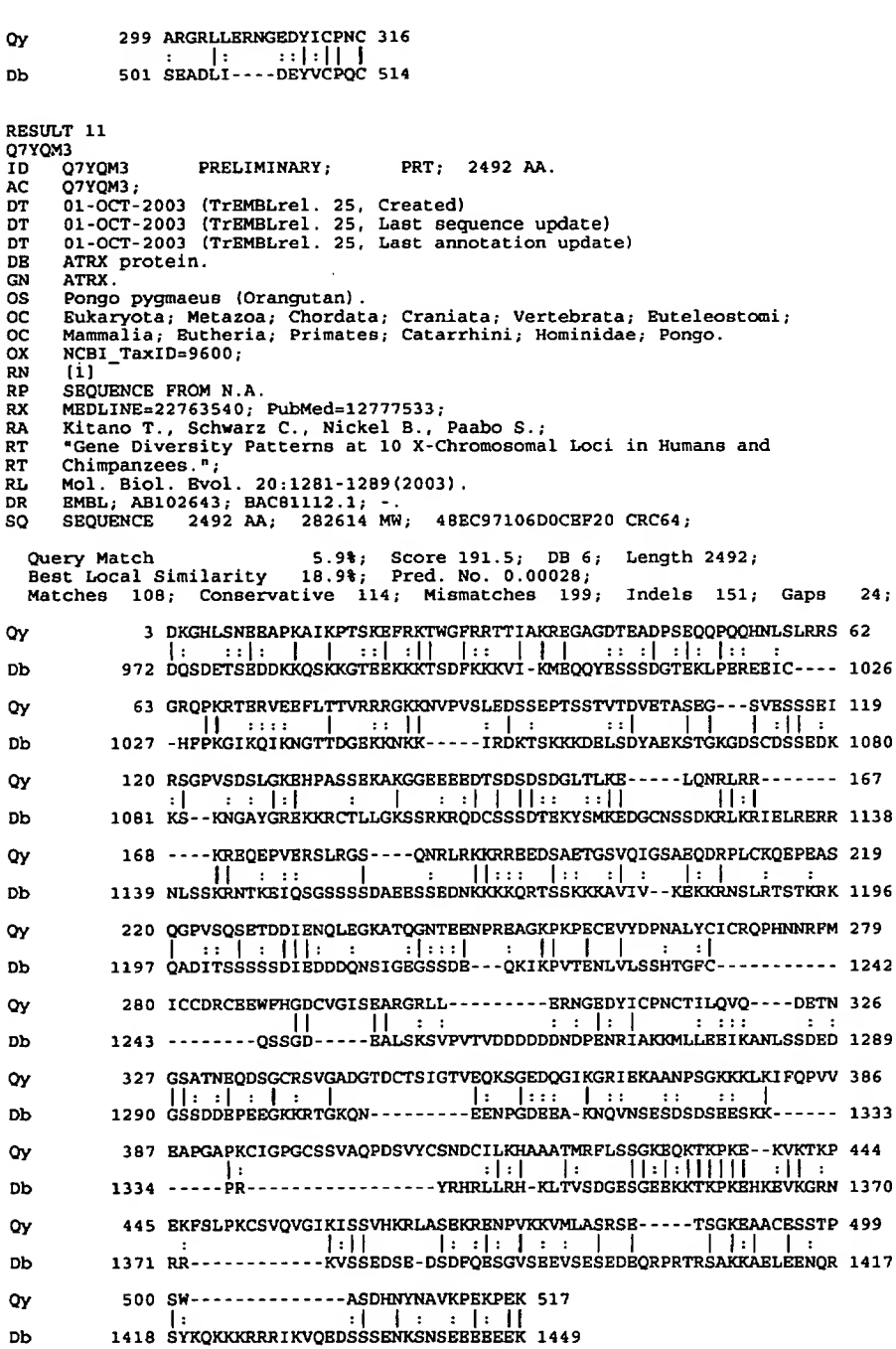
```
RESULT 7
Q9VUB5
ID   Q9VUB5      PRELIMINARY;      PRT;  3146 AA.
AC   Q9VUB5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG9007 protein.
GN   CG9007.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 SET DOMAIN.
DR   EMBL; AE003536; AAF49773.3; -.
DR   FlyBase; FBgn0036398; CG9007.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR001214; SET.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF00628; PHD; 1.
DR   Pfam; PF00856; SET; 1.
DR   SMART; SM00249; PHD; 1.
DR   SMART; SM00317; SET; 1.
DR   PROSITE; PS50280; SET; 1.
DR   PROSITE; PS01359; ZF_PHD_1; 1.
DR   PROSITE; PS50016; ZF_PHD_2; 1.
SQ   SEQUENCE   3146 AA;  336443 MW;  81EEAC5C308FEC31 CRC64;

  Query Match             6.2%;  Score 200.5;  DB 5;  Length 3146;
  Best Local Similarity   21.9%;  Pred. No. 9e-05;
  Matches   93;  Conservative   58;  Mismatches  142;  Indels  131;  Gaps   18;

Qy         86 NVPVSLEDSSEPTSSTVTDVETASEGSVESSSSEIRSGPVSDSLGKEHPASSEKAKGGEEE 145
              |||| :| :: |:|  :          |: ||:  : |      : ||    :    :||
Db        671 NVPVHVE-AAPPSSPAL----------VKGSSQPPAQP-QQQQQQAHPLGPNQLNANDEE 718
```

```
Qy        146 EDTSDSDSDGLTLKE------LQNRLRRKREQEPVERSLRGSQNRLRKKRREEDSAETGS 199
                  | ::|      |   || ::    ::     ||    :::::   |   :
Db        719 ----------LYIEEVRPVPVLTQDLRLQQLHAIMQDHTYASQ----QQQQQPQQAAGDT 764

Qy        200 VQIGSAEQDRPLCKQEPE------------ASQG--------------PVSQSET-DDIE 232
                |:|:|      |:|:             |||                :::|:  ||
Db        765 TNPGAAQQ-----VQQPQQWSLGGIGVTVSGSQGTPTAVGGYCSYFGQQIARSQADDDAH 819

Qy        233 NQLEGKATQG--NTBENPREAGKPKPECEVYDPNALYCICRQPHNNRFMICCDRCEEWFH 290
              : :   :  |  :|: :| |  :  || |  | :   |||   |:: :|||||:|  | |
Db        820 SAISSSSRMGLASTDIDPGEETETAPEABAEDDSVTRCICELTHDDGYMICCDKCSAWQH 879

Qy        291 GDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQD------------SGC 337
              ||:||        :   |:|:|  |    | |:    |   |               :|
Db        880 VDCMGIDR------QNIPEEYMCELCQPRAV-DKARARALQRQKRKEHMLLVATQAANGA 932

Qy        338 RSVGADGTDCTSIGTVEQKSGEDQ-----GIKGRI--------------EKAANPSGKKK 378
              :| | |    :|:   | | |     |: |                | : :  | ||
Db        933 AAVAAGTTLSGGLGSGLPMSEELQHRLASGLNGGFATGTGMSKKSKKTKENSGSTSTLKK 992

Qy        379 LKIFQPVVEAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAAATMRFLSSGKEQKTKPKE 438
              |  :  |   |     | |  :                          ||||  |   |
Db        993 TK--KSAVGMGGEKNASGSGTPTG-----------------------SSGKTSKKSSKR 1026

Qy        439 KVKT 442
              | |:
Db       1027 KSKS 1030
```

```
RESULT 8
Q8T3Y1
ID   Q8T3Y1      PRELIMINARY;      PRT;   366 AA.
AC   Q8T3Y1; Q9W353;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   AT26187p (CG17440 protein).
GN   CG17440.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY089443; AAL90181.1; -.
DR   EMBL; AE003446; AAF46482.2; -.
DR   FlyBase; FBgn0030120; CG17440.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF00628; PHD; 1.
DR   SMART; SM00249; PHD; 1.
DR   PROSITE; PS01359; ZF_PHD_1; 1.
DR   PROSITE; PS50016; ZF_PHD_2; 1.
SQ   SEQUENCE   366 AA;  42048 MW;  CB83AAFEAC514CF3 CRC64;

  Query Match             6.2%;  Score 200;  DB 5;  Length 366;
  Best Local Similarity   28.3%;  Pred. No. 6.7e-06;
  Matches   54;  Conservative   22;  Mismatches   65;  Indels   50;  Gaps    7;
```

```
Qy        267 YCICRQPHNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNC--------TI 318
              |||||    :|||| || ||||:||||: |:|     :    ::| | |          ||
Db         40 YCICRSSDCSRFMIGCDGCEEWYHGDCIEITEKDAEHI----KNYYCRRCKKENPELQTI 95

Qy        319 LQVQDETNGSATNEQD----------SGCRSVGADGTDCT--------SIGTVEQKSGED 360
               ::      :|:|                ||     |     |        :    ::|:
Db         96 FRLVATERAAASNAASTSLNAPGVGPSGAAPAAAPVAPATTSQQAPPPTTAAAKRKNSSA 155

Qy        361 QGIKGRIEKAANPSGKKKLK------------IPQPVV----EAPGAPKCIGPGCSSVA 403
              | |   |    :| |: |                | :    |  |  :| || | | |
Db        156 QEPK---ESQPTQAGTKRDKAAPKTSNVQVSPRAVSPEIFLNPELQGIQQCHGPNCCSHA 212

Qy        404 QPDSVYCSNDC 414
              :| | |||::|
Db        213 RPQSKYCSDEC 223


RESULT 9
O74508
ID   O74508      PRELIMINARY;      PRT;   424 AA.
AC   O74508;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   SPCC594.05c.
GN   SPCC594.05C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972h-;
RA   Rieger M., Lyne M., Rajandream M.A., Barrell B.G.;
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL031523; CAA20664.1; -.
DR   PIR; T41449; T41449.
DR   GeneDB_SPombe; SPCC594.05c; -.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF00628; PHD; 1.
DR   SMART; SM00249; PHD; 1.
DR   PROSITE; PS01359; ZF_PHD_1; 1.
DR   PROSITE; PS50016; ZF_PHD_2; 1.
SQ   SEQUENCE   424 AA;  48683 MW;  2963605C3DFCC0B9 CRC64;

  Query Match             6.1%;  Score 198.5;  DB 3;  Length 424;
  Best Local Similarity   24.1%;  Pred. No. 1e-05;
  Matches   95;  Conservative   46;  Mismatches  139;  Indels  115;  Gaps   18;

Qy         89 VSLEDSSEPTSSTVTD--VETASEGSVESSSEIRSGPVSDSLGKEHPASSEKAKGGEEEE 146
              |  |||:     |:||  :|||:    : :: |        | |       |:  |
Db         20 VKPEDSNR---GTITDPHIETANNEEEKDANVI--------LNKSVKMEVEEVNG---HV 65

Qy        147 DTSDSDSDGLTLKELQNRLRRKREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSAE 206
              |:| :::| : :: :|      |:   ||    ||     || |   :|     : : :|:
Db         66 DSSSTETD-IEMQVIQQPTIPKK--PPVSAHRRGP----RKHRGNANS----QLNLSTAD 114

Qy        207 QDRPLCKQEPEASQGPVSQSETDDIENQLEGKATQGNTEENPREAGKPKPECEVYDPNAL 266
              ||                                                        |
Db        115 HQRP-------------------------------------------------------L 119

Qy        267 YCICRQPHNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETN 326
              ||||::| :  :|: || ||:|||| || | |:    |  :    | || ||     :|
Db        120 YCICQKPDDGSWMLGCDGCEDWFHGTCVNIPESYNDLTVQ----YFCPKCT-----EEGK 170

Qy        327 GSATNEQDSGCRSVGADGTDCTSIGTVEQKSGED---QGIKGRIEKAANPSGKKKLKIPQ 383
              |  | ::   ||          :      | | |   : :|        ||  | | :|
Db        171 GITTWKRK--CRLRECSNPTRPNSNYCSDKHGVDFFREKVK---LSTVEPSAIKNLVLPA 225

Qy        384 PVVEAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAAATMRPLSSGKEQKTKPKEKVKTK 443
```

```
              |      :||   |   |: ||                |::   |  | ||    |
Db        226 KKREEFQNLGTVGPTLPSQVPPEVVY-------NFEIEEANRLNAEIVQLNKEKEVASNK 278

Qy        444 PEKFSLPKCSVQVGIKISSVHKRLASEKRENPVKK 478
                  |  :|: || ||    || ::||  :||
Db        279 -------KIFLQL-IKDSSRRAVLAYKEREG-IKK 304


RESULT 10
Q8VDN7
ID   Q8VDN7      PRELIMINARY;      PRT;   645 AA.
AC   Q8VDN7;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein (Fragment).
GN   FALZ.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC021489; AAH21489.1; -.
DR   MGD; MGI:2444008; Falz.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR001487; Bromodomain.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF00439; bromodomain; 1.
DR   Pfam; PF00628; PHD; 1.
DR   PRINTS; PR00503; BROMODOMAIN.
DR   SMART; SM00297; BROMO; 1.
DR   SMART; SM00249; PHD; 1.
DR   PROSITE; PS00633; BROMODOMAIN_1; 1.
DR   PROSITE; PS50014; BROMODOMAIN_2; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01359; ZF_PHD_1; 1.
DR   PROSITE; PS50016; ZF_PHD_2; 1.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   645 AA;  72167 MW;  6486F34F339820F9 CRC64;

  Query Match             6.1%;  Score 196.5;  DB 11;  Length 645;
  Best Local Similarity   20.1%;  Pred. No. 2.4e-05;
  Matches   76;  Conservative   64;  Mismatches  125;  Indels  113;  Gaps   13;

Qy         43 TEADPSEQQPQQHNLSLRR-------SGRQPKRTERVEEFLTTVRRRGKKNVPVSL-EDS 94
              |  ||  |||   ::: :       |  | :   :::   : | :: |  :|: : ::|
Db        146 TVLSPSCPQPQPQVIAVPQLQQVQVLSQIQSQVVAQIQAQQSGVPQQIKLQLPIQVQQNS 205

Qy         95 SEPTSSTVT-----------------------DVETASEGSVESS--SEIRSGPV--- 124
              :  | | ||                      :||  | ::::|  |||    |
Db        206 AAQTQSVVTVQAASVQEQLQRVQQLRDQQQKKKQQIETEREHTLQASNQSEIIQKQVVMK 265

Qy        125 ---------------------------------SDSLGKEHPASSEKAKGGEE-EEDT 148
                                              | : ||   :::| |  |  |:
Db        266 HNAVIEHLKQKKTMTPAERREENQRMIVCNQVMKYILDKIDKEEKQAAKKRKREESVEQKR 325

Qy        149 SDSDSDGLTL------KELQNRLRRKRE----------QEPVERSLRGSQNRLRKKRREE 192
              |  ::  |:       ::|:  : |||          || ::| |      :::::|
Db        326 SKQNASKLSALLFKHKEQLKABILRKRALLDKELQIQVQEELKRDL-----KMKREREMA 380

Qy        193 DSAETGSVQIGSAEQDRPLCKQEPEASQGPVSQSETDDIENQLEGKAT-------QGNTE 245
              : :  :  : :       |:   | |  |                 | |       |
Db        381 QAVQANAASVPTPSVPAPVPAPAPAAPPAPPRSPPPSTHSLPPAGHPTAPLPVTSQKRKR 440

Qy        246 ENPREAGKPKPEC------EVYDPNALYCICRQPHN-NRFMICCDRCEEWFHGDCVGISE 298
              |  :::   | :      |     |||||: |:: ::| | ||||: |:|| |||| :
Db        441 EEEKDSKSKKKKMISTTSKEAKKDTRLYCICKTPYDESKFYIGCDRCQNWYHGRCVGILQ 500
```



```
Qy        299 ARGRLLERNGEDYICPNC 316
              :   |:    ::|:|| |
Db        501 SEADLI----DEYVCPQC 514


RESULT 11
Q7YQM3
ID   Q7YQM3      PRELIMINARY;      PRT;  2492 AA.
AC   Q7YQM3;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   ATRX protein.
GN   ATRX.
OS   Pongo pygmaeus (Orangutan).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OX   NCBI_TaxID=9600;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22763540; PubMed=12777533;
RA   Kitano T., Schwarz C., Nickel B., Paabo S.;
RT   "Gene Diversity Patterns at 10 X-Chromosomal Loci in Humans and
RT   Chimpanzees.";
RL   Mol. Biol. Evol. 20:1281-1289(2003).
DR   EMBL; AB102643; BAC81112.1; -.
SQ   SEQUENCE   2492 AA;  282614 MW;  48EC97106D0CEF20 CRC64;

  Query Match             5.9%;  Score 191.5;  DB 6;  Length 2492;
  Best Local Similarity   18.9%;  Pred. No. 0.00028;
  Matches  108;  Conservative  114;  Mismatches  199;  Indels  151;  Gaps   24;

Qy          3 DKGHLSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPSEQQPQQHNLSLRRS 62
              |:   ::|:  |  |  ::| :||  |::  | | |   :: :| :|: |::  :
Db        972 DQSDETSEDDKKQSKKGTEEKKKTSDFKKKVI-KMEQQYESSSDGTEKLPEREEIC---- 1026

Qy         63 GRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTVTDVETASEG---SVESSSEI 119
                ||  ::::   |   :: ||    : | :     ::|    | |   | :|| :
Db       1027 -HFPKGIKQIKNGTTDGEKKNKK-----IRDKTSKKKDELSDYAEKSTGKGDSCDSSEDK 1080

Qy        120 RSGPVSDSLGKEHPASSEKAKGGEEEEDTSDSDSDGLTLKE-----LQNRLRR------- 167
              :|   : : |:|    :   |   : :| | ||::  ::||        ||:|
Db       1081 KS--KNGAYGREKKRCTLLGKSSRKRQDCSSSDTEKYSMKEDGCNSSDKRLKRIELRERR 1138

Qy        168 ----KREQEPVERSLRGS----QNRLRKKRREEDSAETGSVQIGSAEQDRPLCKQEPEAS 219
                  ||  : ::     |     :   ||::: |::  :| :   |: |   :  :
Db       1139 NLSSKRNTKEIQSGSSSSDAEESSEDNKKKKQRTSSKKKAVIV--KEKKRNSLRTSTKRK 1196

Qy        220 QGPVSQSETDDIENQLEGKATQGNTEENPREAGKPKPECEVYDPNALYCICRQPHNNRFM 279
              |  :: | : |||:  :    :|:::|   :  ||  |  |   :  :|
Db       1197 QADITSSSSSDIEDDDQNSIGEGSSDE---QKIKPVTENLVLSSHTGFC----------- 1242

Qy        280 ICCDRCEEWFHGDCVGISEARGRLL---------ERNGEDYICPNCTILQVQ----DETN 326
                        ||     ||  : :          : : |: |      : :::      : :
Db       1243 --------QSSGD-----EALSKSVPVTVDDDDDDNDPENRIAKKMLLEEIKANLSSDED 1289

Qy        327 GSATNEQDSGCRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLKIFQPVV 386
              ||: :| : | :  |              |:  |:::  | ::   ::   ::  |
Db       1290 GSSDDEPEEGKKRTGKQN---------EENPGDEEA-KNQVNSESDSDSEESKK------ 1333

Qy        387 EAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAAATMRFLSSGKEQKTKPKE--KVKTKP 444
                |:                    :|:|    |:     ||:|:|||||  :|| :
Db       1334 -----PR-----------------YRHRLLRH-KLTVSDGESGEEKKTKPKEHKEVKGRN 1370

Qy        445 EKFSLPKCSVQVGIKISSVHKRLASEKRENPVKKVMLASRSE-----TSGKEAACESSTP 499
               :           |:||     |: :|: | : :   |  |       | |:|  | :
Db       1371 RR------------KVSSEDSE-DSDPQESGVSEEVSESEDEQRPRTRSAKKAELEENQR 1417

Qy        500 SW--------------ASDHNYNAVKPEKPEK 517
              |:              :|  | :  : |: ||
Db       1418 SYKQKKKRRRIKVQEDSSSENKSNSEEEEEEK 1449


RESULT 12
Q96GI5
ID   Q96GI5      PRELIMINARY;      PRT;  569 AA.
AC   Q96GI5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Muscle;
RA   Strausberg R.;
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC009449; AAH09449.1; -.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   569 AA;  62857 MW;  253856FC544FDBEA CRC64;

  Query Match             5.8%;  Score 188.5;  DB 4;  Length 569;
  Best Local Similarity   21.6%;  Pred. No. 7.3e-05;
  Matches  116;  Conservative   86;  Mismatches  182;  Indels  153;  Gaps   24;

Qy         28 GFRRTT------IAKREGAGDTEAD-----PSEQQPQQHNLSLRRSGRQPKRTERVEEFL 76
              | :| |       ::||      :: |      |||::  : :    ::  |    |:  :
Db         52 GLKRKTPALKMSVSKRARKASSDLDQASVSPSEEENSESSSESEKTSDQDFTPEK--KAA 109

Qy         77 TTVRRRG------KKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDSLGK 130
                 |||       ||  | : :  |: |          |   |  |||    |   ||   |
Db        110 VRAPRRGPLGGRKKKKAPSASDSDSKADSDGAKPEPVAMARSASSSSSSSSSSDSDVSVK 169

Qy        131 EHPASSE-------KAKGGEEEED------TSDSDSDGL-TLKELQNR---------LRR 167
              : |   :       | :| : : :      :||||||  :  : | :  |          ||
Db        170 KPPRGRKPAEKPLPKPRGRKPKPERPPSSSSSDSDSDEVDRISEWKRRDEARRRELEARR 229

Qy        168 KREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSA--------EQDRPLCKQEPEA- 218
              :||||   | ||  :   :::|||   |: |  : ||          | | |: |:  :
Db        230 RREQEEELRRLREQEKEEKERRRER--ADRGEAERGSGGSSGDELREDDEPVKKRGRKGR 287

Qy        219 SQGPVSQSETDDIENQLEGKATQGNTEENPREAGK--PKPECEVYDPNALYCICRQPHNN 276
              :|| | |:::  | :||           ||| |   ||:    :|       |:|
Db        288 GRGPPSSSDSEP-EAELE-----------REAKKSAKKPQSSSTEP------ARKP--- 325

Qy        277 RFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQDSG 336
                                 | | | |                        |:
Db        326 ------------------GQKEKRVR----------------------------PEEKQQ 339

Qy        337 CRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLK------IFQPVVEAPG 390
              : |  : |          :|  |   :  ::||   || ::||:       |   |::|
Db        340 AKPVKVERT---------RKRSEGFSMDRKVEKKKEPSVEEKLQKLHSEIKFALKVDSPD 390

Qy        391 APKCIGPGCSSVAQPDSVYCSNDCILKH--AAATMRFLSSGKEQKTKPKEKVKTKPEKFS 448
               :|:    ::: :  ::  ::  : |:     ||::     | ::|  :  |  |  :
Db        391 VKRCL----NALEELGTLQVTSQILQKNTDVVATLK-----KIRRYKANKDVMEKAAEVY 441

Qy        449 LPKCSVQVGIKISSVHK-RLASEKRENPVKKVMLASRSETSGKEAACESSTPSWASD 504
               |  :| || :| |   |  ::|   :|:   :  | :|:||  | :    ::|
Db        442 TRLKSRVLGPKIEAVQKVNKAGMEKEKAEEKL---AGEELAGEEAPQEKAEDKPSTD 495


RESULT 13
Q9BW08
ID   Q9BW08      PRELIMINARY;      PRT;  670 AA.
AC   Q9BW08;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Similar to hepatoma-derived growth factor, related protein 2.
OS   Homo sapiens (Human).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RA   Strausberg R.;
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC000755; AAH00755.1; -.
DR   InterPro; IPR000313; PWWP_domain.
DR   Pfam; PF00855; PWWP; 1.
DR   SMART; SM00293; PWWP; 1.
DR   PROSITE; PS50812; PWWP; 1.
SQ   SEQUENCE   670 AA;  74229 MW;  BAA24291446BDBAD CRC64;

  Query Match             5.8%;  Score 188.5;  DB 4;  Length 670;
  Best Local Similarity   21.6%;  Pred. No. 9e-05;
  Matches  116;  Conservative   86;  Mismatches  182;  Indels  153;  Gaps   24;

Qy         28 GFRRTT------IAKREGAGDTEAD-----PSEQQPQQHNLSLRRSGRQPKRTERVEEFL 76
              | :| |      ::||     :: |     |||::  : :    ::  |     |:  :
Db        154 GLKRKTPALKMSVSKRARKASSDLDQASVSPSEEENSESSSESEKTSDQDFTPEK--KAA 211

Qy         77 TTVRRRG------KKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDSLGK 130
                    |||       || |  : :   |:  |         |   |  |||   |  ||   |
Db        212 VRAPRRGPLGGRKKKKAPSASDSDSKADSDGAKPEPVAMARSASSSSSSSSSDSDVSVK 271

Qy        131 EHPASSE-------KAKGGEEEED------TSDSDSDGL-TLKELQNR---------LRR 167
              : |    :          | :| : : :          :|||||| :   : | : |          ||
Db        272 KPPRGRKPAEKPLPKPRGRKPKPERPPSSSSSDSDSDEVDRISEWKRRDEARRRELEARR 331

Qy        168 KREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSA--------EQDRPLCKQEPEA- 218
              :||||   | ||   :    :::|||   |:  |   :  ||         | | |: |:  :
Db        332 RREQEEELRRLREQEKEEKERRRER--ADRGEAERGSGGSSGDELREDDEPVKKRGRKGR 389

Qy        219 SQGPVSQSETDDIENQLEGKATQGNTEENPREAGK--PKPECEVYDPNALYCICRQPHNN 276
               :|| | |:::   | :||                 ||| |     ||:     :|        |:|
Db        390 GRGPPSSSDSEP-EAELE------------REAKKSAKKPQSSSTEP------ARKP--- 427

Qy        277 RFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQDSG 336
                                        |  | | |                          |:
Db        428 ------------------GQKEKRVR---------------------------PEEKQQ 441

Qy        337 CRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLK------IFQPVVEAPG 390
               :  |  :  |           :|   |     :   ::||   || ::||:         |   |::|
Db        442 AKPVKVERT---------RKRSEGFSMDRKVEKKKEPSVEEKLQKLHSEIKFALKVDSPD 492

Qy        391 APKCIGPGCSSVAQPDSVYCSNDCILKH--AAATMRFLSSGKEQKTKPKEKVKTKPEKFS 448
                :|:    ::: :  ::  ::  ::  : |:     ||::     | :: |  : |  |  :
Db        493 VKRCL----NALEELGTLQVTSQILQKNTDVVATLK-----KIRRYKANKDVMEKAAEVY 543

Qy        449 LPKCSVQVGIKISSVHK-RLASEKRENPVKKVMLASRSETSGKEAACESSTPSWASD 504
                  |  :| || :| |   |  ::|   :|:   :  | :|:|| | :    ::|
Db        544 TRLKSRVLGPKIEAVQKVNKAGMEKEKAEEKL---AGEELAGEEAPQEKAEDKPSTD 597

RESULT 14
Q7Z4V5
ID   Q7Z4V5      PRELIMINARY;      PRT;   671 AA.
AC   Q7Z4V5;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hepatoma-derived growth factor 2.
GN   HDGF2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yu L.;
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF294267; AAP97281.1; -.
SQ   SEQUENCE   671 AA;  74316 MW;  66CA046E0E802741 CRC64;

  Query Match             5.8%;  Score 188.5;  DB 4;  Length 671;
  Best Local Similarity   21.6%;  Pred. No. 9e-05;
  Matches  116;  Conservative   86;  Mismatches  182;  Indels  153;  Gaps   24;

Qy         28 GFRRTT------IAKREGAGDTEAD-----PSEQQPQQHNLSLRRSGRQPKRTERVEEFL 76
              | :| |      ::||     :: |     |||::  : :    ::  |     |:  :
Db        154 GLKRKTPALKMSVSKRARKASSDLDQASVSPSEEENSESSSESEKTSDQDFTPEK--KAA 211

Qy         77 TTVRRRG------KKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDSLGK 130
                    |||       || |  : :   |:  |         |   |  |||   |  ||   |
Db        212 VRAPRRGPLGGRKKKKAPSASDSDSKADSDGAKPEPVAMARSASSSSSSSSSDSDVSVK 271

Qy        131 EHPASSE-------KAKGGEEEED------TSDSDSDGL-TLKELQNR---------LRR 167
              : |    :          | :| : : :          :|||||| :   : | : |          ||
Db        272 KPPRGRKPAEKPLPKPRGRKPKPERPPSSSSSDSDSDEVDRISEWKRRDEARRRELEARR 331

Qy        168 KREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSA--------EQDRPLCKQEPEA- 218
              :||||   | ||   :    :::|||   |:  |   :  ||         | | |: |:  :
Db        332 RREQEEELRRLREQEKEEKERRRER--ADRGEAERGSGGSSGDELREDDEPVKKRGRKGR 389

Qy        219 SQGPVSQSETDDIENQLEGKATQGNTEENPREAGK--PKPECEVYDPNALYCICRQPHNN 276
               :|| | |:::   | :||                 ||| |     ||:     :|        |:|
Db        390 GRGPPSSSDSEP-EAELE------------REAKKSAKKPQSSSTEP------ARKP--- 427

Qy        277 RFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQDSG 336
                                        |  | | |                          |:
Db        428 ------------------GQKEKRVR---------------------------PEEKQQ 441

Qy        337 CRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLK------IFQPVVEAPG 390
               :  |  :  |           :|   |     :   ::||   || ::||:         |   |::|
Db        442 AKPVKVERT---------RKRSEGFSMDRKVEKKKEPSVEEKLQKLHSEIKFALKVDSPD 492

Qy        391 APKCIGPGCSSVAQPDSVYCSNDCILKH--AAATMRFLSSGKEQKTKPKEKVKTKPEKFS 448
                :|:    ::: :  ::  ::  ::  : |:     ||::     | :: |  : |  |  :
Db        493 VKRCL----NALEELGTLQVTSQILQKNTDVVATLK-----KIRRYKANKDVMEKAAEVY 543

Qy        449 LPKCSVQVGIKISSVHK-RLASEKRENPVKKVMLASRSETSGKEAACESSTPSWASD 504
                  |  :| || :| |   |  ::|   :|:   :  | :|:|| | :    ::|
Db        544 TRLKSRVLGPKIEAVQKVNKAGMEKEKAEEKL---AGEELAGEEAPQEKAEDKPSTD 597

RESULT 15
Q7Z7D6
ID   Q7Z7D6      PRELIMINARY;      PRT;  2764 AA.
AC   Q7Z7D6;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Bromodomain PHD finger transcription factor.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Barak O., Lazzaro M., Lane W., Speicher D., Picketts D.,
RA   Shiekhattar R.;
RT   "Isolation of human NURF.";
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY282495; AAP22284.1; -.
SQ   SEQUENCE   2764 AA;  308079 MW;  9785D7097C2099F9 CRC64;

  Query Match             5.8%;  Score 188.5;  DB 4;  Length 2764;
  Best Local Similarity   19.3%;  Pred. No. 0.00052;
  Matches   70;  Conservative   67;  Mismatches  131;  Indels   95;  Gaps   13;

Qy         31 RTTIAKREGAGDTEADPSEQQPQQHNLSLRR------SGRQPKRTERVEEFLTTVRRRGK 84
              :||::        :|| |:  |   :  :: :               |: | |  :    :  | :|
Db       2289 QTTVSSHV---PSEAQPTHAQSSKPQVAAQSQPQSNVQGQSPVRVQSPSQ--TRIRPSTP 2343
```

```
Qy         85 KNVPVSLEDSSEPTSSTVTDVETASEGSVESSSEIRSGP---------------------- 123
              :    :   : |:|    ::  :   : |  : :| |
Db       2344 SQLSPGQQSQVQTTTSQPIPIQPHTSLQIPSQGQPQSQPQVVMKHNAVIEHLKQKKSMTP 2403

Qy        124 ------------------VSDSLGKEHPASSEKAKGGEE-EEDTSDSDSDGLTL------ 158
                               : | : ||   :::| | |  |:  |  ::  |:
Db       2404 AEREENQRMIVCNQVMKYILDKIDKEEKQAAKKRKREESVEQKRSKQNATKLSALLFKHK 2463

Qy        159 KELQNRLRRKRE----------QEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSAEQD 208
              ::|:  : :||           || ::| |     :::|::     |:  :|     |
Db       2464 EQLRAEILKKRALLDKDLQIEVQEELKRDL-----KIKKEKDLMQLAQATAV----AAPC 2514

Qy        209 RPLCKQEPEASQGPVSQSETDDIEN-------QLEGKATQGNTEENPREAGKPKPE---- 257
               |:    |      | |      :::         |   : :   ||    : | | :
Db       2515 PPVTPAPPAPPAPPPSPPPPPAVQHTGLLSTPTLPAASQKRKREEEKDSSSKSKKKKMIS 2574

Qy        258 ---CEVYDPNALYCICRQPHN-NRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYIC 313
                 |      |||||: |:: ::| | |||| |:|| |||| ::   |:    ::|:|
Db       2575 TTSKETKKDTKLYCICKTPYDESKFYIGCDRCQNWYHGRCVGILQSEAELI----DEYVC 2630

Qy        314 PNC 316
              | |
Db       2631 PQC 2633
```

Search completed: April 20, 2004, 21:19:50
Job time : 48 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 20, 2004, 21:13:09 ; Search time 20 Seconds
(without alignments)
2953.080 Million cell updates/sec

Title: US-09-787-016A-4
Perfect score: 3228
Sequence: 1 MDDKGHLSNEEAPKAIKPTS.....LKSCVGLMLEAISYFSFRPW 614

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	198.5	6.1	424	2	T41449	probable phd finge
2	197.5	6.1	1805	2	A34736	nestin - rat
3	185	5.7	1359	2	T34036	hypothetical prote
4	184.5	5.7	1641	2	I38614	helicase II - huma
5	179	5.5	385	2	T19201	hypothetical prote
6	177	5.5	452	2	T21435	hypothetical prote
7	177	5.5	1110	2	I51116	NP-180 - sea lampr
8	172.5	5.3	1684	2	JW0057	gravin - human
9	170.5	5.3	405	2	T21433	hypothetical prote
10	170.5	5.3	489	2	A45988	dentin matrix acid
11	167	5.2	5327	2	T13564	microtubule-associ
12	165	5.1	6642	2	T29757	protein UNC-89 - C
13	164.5	5.1	510	2	T21430	hypothetical prote
14	163.5	5.1	990	2	I51618	nucleolar phosphop
15	161.5	5.0	952	2	T18837	hypothetical prote
16	161	5.0	3488	2	T34418	hypothetical prote
17	160.5	5.0	852	2	T06310	hypothetical prote
18	160	5.0	1004	2	A55142	myosin-light-chain
19	159.5	4.9	706	2	A45990	junctional sarcopl
20	159	4.9	669	2	JC5662	hepatoma-derived g
21	158.5	4.9	751	2	S38101	hypothetical prote
22	156.5	4.8	522	2	C96608	hypothetical prote
23	156.5	4.8	611	2	T22456	hypothetical prote
24	156.5	4.8	3507	2	T34513	hypothetical prote
25	156	4.8	763	2	T08929	hypothetical prote
26	155	4.8	699	2	T01029	hypothetical prote
27	155	4.8	729	2	S68191	triadin - human
28	155	4.8	2275	2	T33123	hypothetical prote
29	154.5	4.8	734	2	B42680	nucleolus-cytoplas
30	154.5	4.8	1151	2	T24541	hypothetical prote
31	154.5	4.8	1560	2	T30282	calcium-binding pr
32	154	4.8	555	2	T26413	hypothetical prote
33	154	4.8	1390	2	T14004	trfA protein - sli
34	153.5	4.8	1849	2	C41859	IgA-specific metal
35	153	4.7	1165	2	T16420	hypothetical prote
36	151	4.7	2282	2	T42717	DNA-binding protei
37	150.5	4.7	1077	2	T21800	hypothetical prote
38	150	4.6	1280	2	G96796	hypothetical prote
39	148	4.6	657	2	S10001	MyD116 protein - m
40	148	4.6	863	2	T49709	related to glucan
41	147.5	4.6	258	2	T06070	ES43 protein homol
42	147.5	4.6	258	2	H85462	ES43 like protein
43	147.5	4.6	1166	2	H86341	hypothetical prote
44	147	4.6	1507	2	B47328	natural killer cel
45	146.5	4.5	644	2	S55395	neurofilament prot

ALIGNMENTS

RESULT 1
T41449
probable phd finger transcription regulator - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 07-Dec-1999
C;Accession: T41449
R;Rieger, M.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z21994
A;Accession: T41449
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-424 <RIE>
A;Cross-references: EMBL:AL031523; PIDN:CAA20664.1; GSPDB:GN00068; SPDB:SPCC594.05c
A;Experimental source: strain 972h-; cosmid c594
C;Genetics:
A;Gene: SPDB:SPCC594.05c
A;Map position: 3

Query Match 6.1%; Score 198.5; DB 2; Length 424;
Best Local Similarity 24.1%; Pred. No. 0.00042;
Matches 95; Conservative 46; Mismatches 139; Indels 115; Gaps 18;

```
Qy         89 VSLEDSSEPTSSTVTD--VETASEGSVESSSSEIRSGPVSDSLGKEHPASSEKAKGGEEEE 146
              |  |||:     |:||  :|||:    : :: |         | |       |:  |
Db         20 VKFEDSNR---GTITDFHIETANNEEEKDANVI--------LNKSVKMEVEEVNG---HV 65

Qy        147 DTSDSDSDGLTLKELQNRLRRKREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSAE 206
              |:| :::| : :: :|     |:   ||    ||     || |   :|     : : :|:
Db         66 DSSSTETD-IEMQVIQQPTIPKK--PPVSAHRRGP----RKHRGNANS----QLNLSTAD 114

Qy        207 QDRPLCKQEPEASQGPVSQSETDDIENQLEGKATQGNTEENPREAGKPKPECEVYDPNAL 266
              ||                                                         |
Db        115 HQRP-----------------------------------------------------L 119

Qy        267 YCICRQPHNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETN 326
              |||::| :  :|: || ||:|||| || | |:   |  :    | || ||     :|
Db        120 YCICQKPDDGSWMLGCDGCEDWFHGTCVNIPESYNDLTVQ----YFCPKCT-----EEGK 170

Qy        327 GSATNEQDSGCRSVGADGTDCTSIGTVEQKSGED---QGIKGRIEKAANPSGKKKLKIFQ 383
              |  | ::   ||          :     | | |   : :|       ||  | | :|
Db        171 GITTWKRK--CRLRECSNPTRPNSNYCSDKHGVDFFREKVK---LSTVEPSAIKNLVLFA 225

Qy        384 PVVEAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAAATMRPLSSGKEQKTKPKEKVKTK 443
                    :||   |  |: ||                  |::   | | ||    |
Db        226 KKREEFQNLGTVGPTLPSQVPPEVVY-------NFEIEEANRLNAEIVQLNKEKEVASNK 278

Qy        444 PEKPSLPKCSVQVGIKISSVHKRLASEKRENPVKK 478
                  | :|: || ||    || ::||  :||
Db        279 -------KIFLQL-IKDSSRRAVLAYKEREG-IKK 304
```

RESULT 2
A34736
nestin - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 13-Jul-1990 #sequence_revision 09-Oct-1992 #text_change 05-Nov-1999
C;Accession: A34736
R;Lendahl, U.; Zimmerman, L.B.; McKay, R.D.G.
Cell 60, 585-595, 1990
A;Title: CNS stem cells express a new class of intermediate filament protein.
A;Reference number: A34736; MUID:90150286; PMID:1689217
A;Accession: A34736
A;Status: preliminary
A;Molecule type: DNA; mRNA
A;Residues: 1-1805 <LEN>
A;Cross-references: GB:M34384; NID:g205663; PIDN:AAA41685.1; PID:g205664
A;Note: the authors translated the codon CGG for residue 168 as Gly, ATC for residue 169
C;Keywords: coiled coil

Query Match 6.1%; Score 197.5; DB 2; Length 1805;
Best Local Similarity 21.2%; Pred. No. 0.0023;
Matches 143; Conservative 99; Mismatches 223; Indels 211; Gaps 35;

```
Qy          2 DDKG---HLSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPSEQQPQQHNLS 58
              :|:|    |  :||   :|  :   :||          :|  |||    | | |   :
Db        486 EDEGQIWELVEKEADIEVKVENSSAQKT----------QESGLDTEETQDSQGPLQKE-T 534

Qy         59 LRRSGRQPKRTERVEEFLTTVRRRGKKNVPVSLEDS----SEPTSSTVTDVETASEGSVE 114
               |:  | :|  : ::: : |     ||:|   | |        |       :::  | :||
Db        535 LKALGEEPLMSLKIQNYETA----GKENCNSSTEGHLGTLEGPEKEKQIPLKSLEEKNVE 590

Qy        115 SSSEIRSG-PV-SDSLGKEHPASSEK----AKG--------GEEEEDTSDSDSDG----- 155
              |   : :| || |: ||||   : ::    ||         |:| ::       :|
Db        591 SEKTLENGVPVLSELLGKEDTRTEDQELMSPKGTLKRFSSLGKESQEVVRPSKEGNLESW 650

Qy        156 ----------------------LTLKELQN-------------RLRRKREQEP------ 173
                                   |  || |:             |  :|  |||
Db        651 TAFKEESQHPLGFPGAEDQMLERLVEKEDQSFPRSPEEEDQEACRPLQKENQEPLGYEEA 710

Qy        174 ----VERSL-RGSQNRLRKKRREEDSAETGSVQIGSAEQDRPLCKQEPEASQGPVSQSET 228
                  :|| : : ||  ||    ||:  | |          | | |:    :| |:   |
Db        711 EGQILERLIEKESQESLRSP--EEEDQEAG----------RSLQKE----NQEPLGYEEA 754

Qy        229 DD--IENQLEGKATQG--NTEENPREAGKP-----KPECEVYDPNALYCICRQPHNNRFM 279
              :|  :|  :| :: :  : ||| |  |||     :      : |   :  : | | :
Db        755 EDQMLERLIEKESQESLKSPEENQR-IGKPLERENQKSLRYLEENQETFVPLESRNQRPL 813

Qy        280 ICCDRCEE----------------------------WFHGDCVGIS-------EARGRL 303
                :  ||                            ||:  |       :: | |
Db        814 RSLEVEEEEQRIVKPLEKVSQDSLGSLAEENVQPLRYLEEDDCINKSLLEDKTHKSLGSL 873

Qy        304 LERNGEDYICPNCTILQV----------------QDETNGSATNEQDSGCRSVGADGTD 346
                :|||:  | |  :  ||                ::    | ::|::        :| :
Db        874 EDRNGDSIIIPQESETQVSLRPPEEEDQRIVNHLEKESQEFSRSSEEEEQVMERSLEGEN 933

Qy        347 CTSIGTVEQKSGEDQGIKGRIEKAANPSGK----KKLKIFQPVVEAPGAPKCIGPGCSSV 402
              |: :||:   ||| :: ::|| :  |||    :  : | |:       |          |:
Db        934 HESLSSVEK---EDQMVESQLEKESQDSGKSLEDESQETFGPL------EKENAESLRSL 984

Qy        403 AQPDSVYCSNDCILKHAAATMRFLSSGKEQ-KTKPKEKVKTKPEKPSLPKCSVQVGIKIS 461
              | |        : :    |:|  : | ||    | |||   ||   |||    :|
Db        985 AGQDQ---EEQKLEQETQQTLR--AVGNEQMAVSPPEKV--DPE---LPK---PLG-NDQ 1030

Qy        462 SVHKRLASEKREN----------PVKKV------MLASRSETSGKEAACESSTPSWASDH 505
              : : |  | :|:           || :      | :  |   :  : :: : | |:::
Db       1031 EIARSLGKENQESLVSLKEKGIETVKSLETEIIEPLETAEEDLERRKSIDTQEPLWSTEV 1090

Qy        506 NYNAVKPEKPEKPTAL 521
                |:| : | | :|
Db       1091 ARETVEPPEDEPPGSL 1106
```

RESULT 3
T34036

hypothetical protein B0041.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34036
R;Fulton, R.; Wohldmann, P.
submitted to the EMBL Data Library, April 1997
A;Description: The sequence of C. elegans cosmid B0041.
A;Reference number: Z21466
A;Accession: T34036
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1359 <FUL>
A;Cross-references: EMBL:AF000196; PIDN:AAC24256.1; GSPDB:GN00019; CESP:B0041.7
A;Experimental source: strain Bristol N2; clone B0041
C;Genetics:
A;Gene: CESP:B0041.7
A;Map position: 1
A;Introns: 12/2; 59/2; 248/2; 582/1; 675/1; 733/3; 966/2; 1044/2; 1310/3

Query Match 5.7%; Score 185; DB 2; Length 1359;
Best Local Similarity 21.4%; Pred. No. 0.0086;
Matches 146; Conservative 88; Mismatches 217; Indels 232; Gaps 33;

```
Qy          8 SNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPSE--QQPQQHNLSLRRSGRQ 65
              |  | | :| : |: :|        : |:|   : ||   :  :: ::   |  |
Db        210 SESEDEKEVKKSKKKSKK--------VVKKESESEDEAPEKKKTEKRKRSKTSSEESSES 261

Qy         66 PKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTVTDVETASE----GSV----ESSS 117
              |  |  ||  :: : :  ||  |::::  |     |  :|||   :    |:|    :|
Db        262 EKSDEEEEEKESSPKPKKKK--PLAVKKLSSDEESEESDVEVLPQKKKRGAVTLISDSED 319

Qy        118 EIRSGPVSDSLGKEHPASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLRRKREQ-EPVER 176
              |      |::   |  | :|||  |  |  |||    :|:    ||  :|:|: |  ::
Db        320 EKDQKSESEASDVEEKVSKKKAKKQESSESGSDSSEGSITV----NRKSKKKEKPEKKKK 375

Qy        177 SLRGSQNRLRKKRREEDSAETGSVQIGSAEQDRPLCKQEPEASQGPVSQSETDDIENQLE 236
              :    ::|:|:  : : ||          |: : | |:: | :   :   | :|:   |
Db        376 GIIMDSSKLQKETIDAERAE--------KERRKRLEKKQKEPN--GIVLEEGEDLTEMLT 425

Qy        237 GKATQGNTEE---NP------REAGKPKPECEVYDPNALYCICR--QPHNNRFMICC--- 282
              | ::|   :   :|       |: ||    ||:  |:|  | :  | |  :|| |
Db        426 GTSSQRKLKSVVLDPDSSTVDEESKKP---VEVH--NSLVRILKPHQAHGIQFMYDCAPE 480

Qy        283 --DRCEEWFHG----DCVGISEARGRLL------------ERNGED-----YICPNCTIL 319
                || :    |    |:|:   |: |             |: ||       : |   |:
Db        481 SLDRLDTEGSGGILAHCMGL----GKTLQVITFLHTVLMHEKIGEKCKRVLVVVPKNVII 536

Qy        320 ----------QVQDETNGSATNEQD----------------SGCRSVGADGTDCTSIG 351
                       :| :   || |                  |   ||  | | |
Db        537 NWPKEFQKWLVDNDEELDTIDVNELDSYKTIEDRRRALKAWHSSKTPSVMIIGYDLFRIL 596

Qy        352 TVEQ--KSGEDQGIKGRIEKAANPSGKKKLKIFQPVVEAPGAPKCIGPGCSSVAQPDSVY 409
              |||   |  : :    |:|||            |:  :: ||              || |
Db        597 TVEDDPKKKKPKNRNRRLEKAKED--------FRKYLQNPG-------------PDMVV 634

Qy        410 CSNDCILKHAAATMRFLSSGKEQKTKPKEKVKTKPEKPSLPKCSVQVGIKISSVHKRLAS 469
              |                        | | | :  :| || |::
Db        635 CD------------------------EAHKLKNDDSALSKCMVKI------------- 655

Qy        470 EKRENPVKKVMLASRSETSGKEAACESSTPSWASDHNYNAVKPEKPEKPTALSPTLLSKC 529
                                  |   | : ||                    |   |:
Db        656 -----------------LTKRRICLTGTP--------------------LQNNLME-- 674

Qy        530 TYHPKAGFPGPSHHLGGCLGLSRTRVLGVLVLIVASSSLPARSRYQDASGPQV-FLPSLW 588
              ||     |  |     | || ::|      | |:       | | :|||  :| |:
Db        675 -YHCMVNFVKP-----GLLG-TKTEFANRFVNII------NRGRTKDASPLEVSFMKRRC 721

Qy        589 SLSGWFLKSCVG-----LMLEAI 606
              :  || ||    ::  |||
Db        722 HVLYDHLKKCVDRKDYRVLTEAI 744
```

RESULT 4

```
I38614
helicase II - human
C;Species: Homo sapiens (man)
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 05-Nov-1999
C;Accession: I38614
R;Stayton, C.L.; Dabovic, B.; Gulisano, M.; Gecz, J.; Broccoli, V.; Giovanazzi, S.; Boss
Hum. Mol. Genet. 3, 1957-1964, 1994
A;Title: Cloning and characterization of a new human Xq13 gene, encoding a putative heli
A;Reference number: I38614; MUID:95179111; PMID:7874112
A;Accession: I38614
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1641 <RES>
A;Cross-references: EMBL:U09820; NID:g606832; PIDN:AAC50069.1; PID:g606833
C;Genetics:
A;Gene: RAD54L

  Query Match             5.7%;  Score 184.5;  DB 2;  Length 1641;
  Best Local Similarity   18.7%;  Pred. No. 0.011;
  Matches 107;  Conservative  113;  Mismatches  201;  Indels  151;  Gaps   24;

Qy          3 DKGHLSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPSEQQPQQHNLSLRRS 62
Db        113 DQSDETSEDDKKQSKKGTEEKKKPSDFKKKVI-KMEQQYESSSDGTEKLPEREEIC---- 167

Qy         63 GRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTVTDVETASEG---SVESSSEI 119
Db        168 -HFPKGIKQIKNGTTDGEKKSKK-----IRDKTSKKKDELSDYAEKSTGKGDSCDSSEDK 221

Qy        120 RSGPVSDSLGKEHPASSEKAKGGEEEEDTSDSDSDGLTLKE-----LQNRLRR------- 167
Db        222 KS--KNGAYGREKKRCKLLGKSSRKRQDCSSSDTEKYSMKEDGCNSSDKRLKRIELRERR 279

Qy        168 ----KREQEPVERSLRGS----QNRLRKKRREEDSAETGSVQIGSAEQDRPLCKQEPEAS 219
Db        280 NLSSKRNTKEIQSGSSSSDAEESSEDNKKKKQRTSSKKKAVIV--KEKKRNSLRTSTKRK 337

Qy        220 QGPVSQSETDDIENQLEGKATQGNTEENPREAGKPKPECEVYDPNALYCICRQPHNNRFM 279
Db        338 QADITSSSSSDIEDDDQNSIGEGSSDE---QKIKPVTENLVLSSHTGFC----------- 383

Qy        280 ICCDRCEEWFHGDCVGISEARGRLL---------ERNGEDYICPNCTILQVQ----DETN 326
Db        384 --------QSSGD-----EALSKSVPVTVDDDDDNDPENRIAKKMLLEEIKANLSSDED 430

Qy        327 GSATNEQDSGCRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLKIFQPVV 386
Db        431 GSSDDEPEEGKKRTGKQN---------EENPGDEEA-KNQVNSESDSDSEESKK------ 474

Qy        387 EAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAAATMRFLSSGKEQKTKPKE--KVKTKP 444
Db        475 -----PR-----------------YRHRLLRH-KLTVSDGESGEEKKTKPKEHKEVKGRN 511

Qy        445 EKFSLPKCSVQVGIKISSVHKRLASEKRENPVKKVMLASRSE-----TSGKEAACESSTP 499
Db        512 RR------------KVSSEDSE-DSDFQESGVSEEVSESEDEQRPRTRSAKKAELEENQR 558

Qy        500 SW--------------ASDHNYNAVKPEKPEK 517
Db        559 SYKQKKKRRRIKVQEDSSSENKSNSEEEEEEK 590

RESULT 5
T19201
hypothetical protein C11G6.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T19201
R;Wilkinson, J.
submitted to the EMBL Data Library, March 1996
A;Reference number: Z19089
A;Accession: T19201
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-385 <WIL>
A;Cross-references: EMBL:Z70204; PIDN:CAA94113.1; GSPDB:GN00028; CESP:C11G6.3
A;Experimental source: clone C11G6
C;Genetics:
A;Gene: CESP:C11G6.3
A;Map position: X
A;Introns: 12/1; 106/3; 171/2; 335/2

  Query Match             5.5%;  Score 179;  DB 2;  Length 385;
  Best Local Similarity   21.9%;  Pred. No. 0.0047;
  Matches   74;  Conservative   56;  Mismatches  148;  Indels   60;  Gaps   12;

Qy          4 KGHLSNEEAPKAIKPTSKEFR-------------KTWGFRRTTIAKREGAG-DTEADPSE 49
Db         54 KSNYHNEPPPPPANPPPLKFRFKNLFNLGDQDVKKEEPSSMTPESSRPGSSLETPSSSSS 113

Qy         50 QQPQQHNLSLRRSGRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTVTDVETAS 109
Db        114 SKHHHHHHKKERKDKEHKKHKKDREHRDREKERERD------ERKERERQQKEKEREDAA 167

Qy        110 EGSVESSSEIRSGPVSDSLGKEHPASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLRRKR 169
Db        168 RREIEEKAEMDAKRVAE---EEEERRKEKEKRREEKKKQKE------LLKEKERSERKEK 218

Qy        170 EQEPVERSLRGSQNRLRKKRRE-------EDSAETGSVQIGSAEQDRPLCKQEPEASQGP 222
Db        219 ERE-LEREKEKSREKEREKEREKEREKEREKEREKQKEREKEREKEREKEKKREEEARRK 277

Qy        223 VSQSETDDIENQLEGKATQGNTEENPREAGKP---KPECEVYDPNALYCICRQPHNNRFM 279
Db        278 KEEASTPVIIRPL---LSQEDDDSNGSESSEEIWICPVCSV-----AYTV------GANM 323

Qy        280 ICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCT 317
Db        324 IGCDQCQDWFHWHCVGLT------AEPTDSKWFCTRCT 355

RESULT 6
T21435
hypothetical protein F26H11.3c - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
C;Accession: T21435
R;Barlow, K.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19421
A;Accession: T21435
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-452 <WIL>
A;Cross-references: EMBL:Z81515; PIDN:CAB54234.1; GSPDB:GN00020; CESP:F26H11.3c
A;Experimental source: clone F26H11
C;Genetics:
A;Gene: CESP:F26H11.3c
A;Map position: 2
A;Introns: 58/1; 89/3; 230/2; 290/2; 334/1
C;Superfamily: bromodomain homology
F;313-368/Domain: bromodomain homology <BRO>

  Query Match             5.5%;  Score 177;  DB 2;  Length 452;
  Best Local Similarity   22.1%;  Pred. No. 0.0072;
  Matches   67;  Conservative   43;  Mismatches  117;  Indels   76;  Gaps   11;

Qy         91 LEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDSLGKEHPASSEKAKGGEEEEDTSD 150
Db         16 ITDHMERLKKEINKRRTRMENEAEQQCGLLT-PWRKARARPHRAAKPKAEVKKEVINPAD 74

Qy        151 SDSDGLTLKELQNRLRRKREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSAEQDRP 210
Db         75 ITLGGDTYDYV-------KEQKPTE-SIATNVSRRRRT----------SANLSKSEDDRD 116

Qy        211 LCKQE---PEASQGPVSQSETDDIENQLEGKATQGN--------TEENPREAGKPKPECE 259
```

```
Db        117 KPESQSTAPKSKERRTSEPPASHVAFHTPGSATPHDINLSIEHCTCQKIFDASKLYIQCE 176

Qy        260 ---------------------------------VYDPNALYCICRQPHNN-RFMICCD 283

Db        177 LCARWYHGDCVGVAEQTILGLEHWSCEECIEEQERVKDQPALYCVCQKPYDDTKFYVGCD 236

Qy        284 RCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQD-SGCRSVGA 342

Db        237 SCQGWFHPECVGTTRAEA----EQAADYNCPACT-----REAEGYESEASDVSGSSRVSV 287

Qy        343 DGT 345

Db        288 QLT 290
```

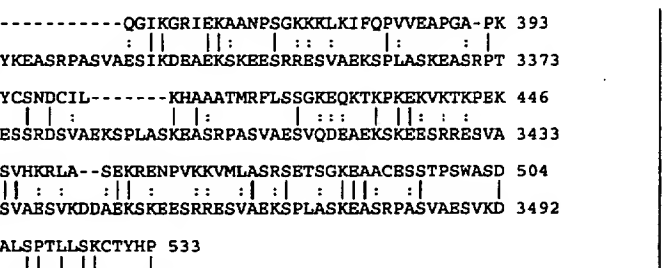
RESULT 7
I51116
NF-180 - sea lamprey
C;Species: Petromyzon marinus (sea lamprey)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: I51116
R;Jacobs, A.J.; Kamholz, J.; Selzer, M.E.
Brain Res. Mol. Brain Res. 29, 43-52, 1995
A;Title: The single lamprey neurofilament subunit (NF-180) lacks multiphosphorylation re
A;Reference number: I51116; MUID:95287814; PMID:7770000
A;Accession: I51116
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1110 <JAC>
A;Cross-references: EMBL:U19361; NID:g632548; PIDN:AAA80106.1; PID:g632549
C;Superfamily: neurofilament triplet H protein

```
  Query Match             5.5%;  Score 177;  DB 2;  Length 1110;
  Best Local Similarity   18.9%;  Pred. No. 0.019;
  Matches  102;  Conservative   74;  Mismatches  197;  Indels  168;  Gaps   17;

Qy          2 DDKGHL-----SNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPSEQQPQ-QH 55

Db        509 EDRGRKEGEAEAEEEAEEEVEKEEAE--------EAEVEEAE-AEETEAEAAEEEEEAEG 559

Qy         56 NLSLRRSGRQPKRTERVEEFLTTVRRRGKKNVPV----------SLEDSSEPTSSTVTDV 105

Db        560 EEEAEAEGEEAEEAEEVEEEAIEKAEAAEAKAEVEEEEAEAEEEEEEEAEEEEVEAETKE 619

Qy        106 ETASEGSVESSSEIRSGPVSDSLGKEHPASSEKAKGGE---EEEDTSDSDSDGLTLKELQ 162

Db        620 EVEAEAEVEEEGEAAEEEAEEEEAEEEEVTSKKAKTQEAEVEEEEAEAAEA------EAE 673

Qy        163 NRLRRKREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSAEQDRPLCKQEPEASQGP 222

Db        674 EEAEEEAGEEDVEAE--------SKEEEEEDSKE------ADAEEDE---AEEEEVKEEE 716

Qy        223 VSQSETDDIENQLEGKATQGNTEENPREAGKPKPECEVYDPNALYCICRQPHNNRFMICC 282

Db        717 VTKSDAEEAEAEAEEEAAK--SEEEAAEEAKDEAEEEEAEEEAV---------------- 758

Qy        283 DRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQDSGCRSVGA 342

Db        759 ----------------------------------------EETEAATEEAEA------- 770

Qy        343 DGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLKIFQPVVEAPGAPKCIGPGCSSV 402

Db        771 ------------KEASDDEKPEEEVKESEAPVAPEAKKAPEP----KAAPKKKAP--AKV 812

Qy        403 AQPDSVYCSNDCILKHAAATMRFLSSGKEQKTKPKEKVKTKPEKPSLPKCSVQVGIKISS 462

Db        813 ESPTS---------------------EPEDEPKAEVVEKKGKAEAPKPKA-------- 841

Qy        463 VHKRLASEKRENPVKKVMLASRSETSGKEAACESSTPSWASDHNYNAVKPEKPEKPTALS 522

Db        842 --KPAAAKKEAKPVEKEEEPEESPTEEEPKKPAAAKPAKAPAKPKPAPKAEAEEKPEPAK 899

Qy        523 P 523

Db        900 P 900
```

RESULT 8
JW0057
gravin - human
C;Species: Homo sapiens (man)
C;Date: 13-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 08-Oct-1999
C;Accession: JW0057
R;Sato, N.; Kokame, K.; Shimokado, K.; Kato, H.; Miyata, T.
J. Biochem. 123, 1119-1126, 1998
A;Title: Changes of gene expression by lysophosphatidylcholine in vascular endothelial ce
A;Reference number: JW0057; MUID:98269042; PMID:9604001
A;Accession: JW0057
A;Molecule type: mRNA
A;Residues: 1-1684 <SAT>
A;Cross-references: DDBJ:AB003476; NID:g2081606; PIDN:BAA19927.1; PID:d1020716; PID:g2081
C;Comment: This protein regulates cell growth.
F;433-439/Region: nuclear location signal
F;522-527/Region: nuclear location signal
F;591-596/Region: nuclear location signal
F;671-676/Region: nuclear location signal

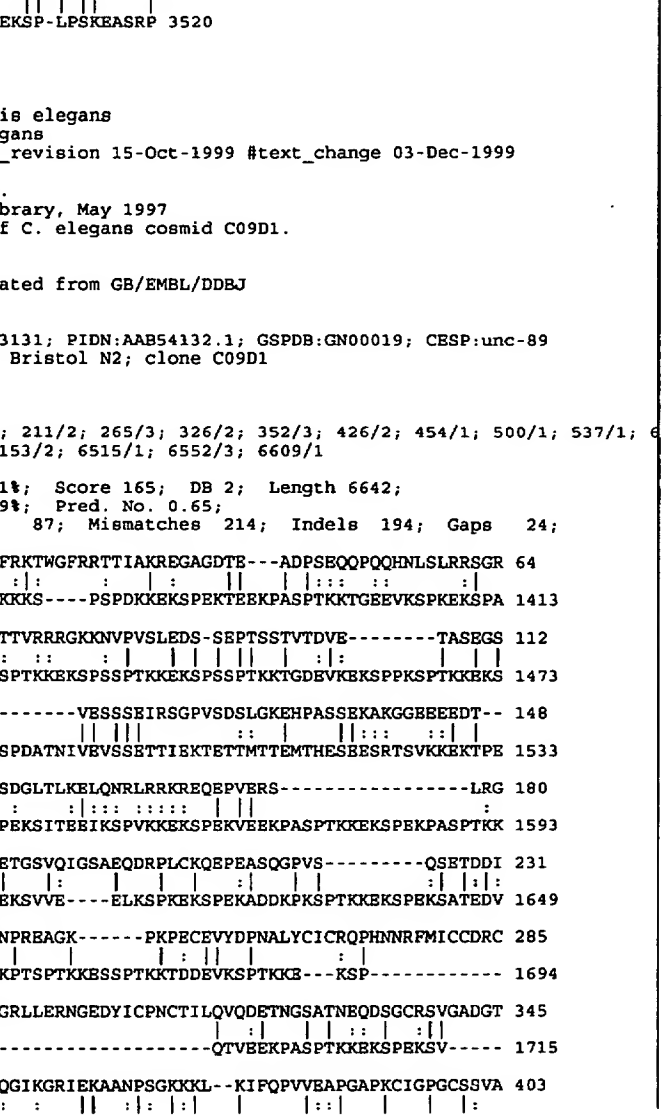
```
  Query Match             5.3%;  Score 172.5;  DB 2;  Length 1684;
  Best Local Similarity   21.4%;  Pred. No. 0.055;
  Matches  135;  Conservative   80;  Mismatches  212;  Indels  203;  Gaps   32;

Qy         10 EEAPKAI--------------KPTSKE-----FRKTW---GFRRT--------------- 32

Db         42 EETPEIIEQIPSSESNLEELTQPTESQANDIGFKKVFKFVGFKFTVKKDKTEKPDTVQLL 101

Qy         33 TIAKREGAGDTEA----DPS---------EQQPQQHNLSLRRSGRQPKRTERVEEFLTTV 79

Db        102 TVKKDEGEGAAGAGDHKDPSLGAGEAASKESEPKQ-------------STEKPEETLKRE 148

Qy         80 RRRGKKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDSLGK--------- 130

Db        149 QSHAEISPPA---ESGQAVEECKEEGEEKQEKEPSKSAESPTSPVTSETGSTFKKFFTQG 205

Qy        131 ----EHPASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLRRKREQEPVERSLRGSQNRLR 186

Db        206 WAGWRKKTSFRKPK--EDEVEASE----------------KKKEQEP------------E 235

Qy        187 KKRREED-SAETGSVQIGSAEQDRPLCKQEP-EASQGPVSQSETDDI----ENQLEGKAT 240

Db        236 KVDTEEDGKAEVASEKLTASEQAHP---QEPAESAHEPRLSAEYEKVELPSEEQVSG--S 290

Qy        241 QGNTEENPREAGKPKP-ECEVYDPNALYCICRQPHNNRFM--ICCDRCEEWFHGDCVGIS 297

Db        291 QGPSEE------KPAPLATEVFDEKI------EVHQEEVVAEVHVSTVEERTEEQKTEVE 338

Qy        298 EARG-----RLLERNGE--------DYICPNCTILQVQDETNGSATNEQDSGCRSVGADG 344

Db        339 ETAGSVPAEELVEMDAEPQEAEPAKELVKLKETCVSGEDPTQG-ADLSPDEKVLSKPPEG 397

Qy        345 TDCTSIGTVEQKSGEDQ-GIKGR-IEKAANPSGKKKLKIFQPVVEAPGAPKCIG------ 396

Db        398 V----VSEVEMLSSQERMKVQGSPLKKLFTSTGLKKLSGKKQKGKRGGGDEESGEHTQVP 453

Qy        397 -------------PGCSSVAQPDSVYCSNDCILKHAAATMRFLSSGKEQKTKPKEKVK-- 441

Db        454 ADSPDSQEEQKGESSASSPEEPEEITC-----LEKGLAEVQQDGEAEEGATSDGEKKREG 508

Qy        442 ----TKPEKPSLPKCSVQVGIKISSVHKRLASEKRENPVKKVMLASRSETSGKEAACESS 497

Db        509 VTPWASFKKMVTPKKRV-----------RRPSESDKEDELDKVKSATLSST--------ES 550

Qy        498 TPSWASDHNYNAVKPEKPEKPTALSPTLLS 527

Db        551 TASEMQEEMKGSVEEPKPEEPKRKVDTSVS 580
```

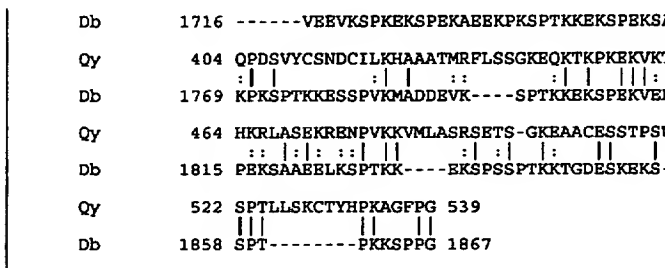
```
RESULT 9
T21433
hypothetical protein F26H11.3a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
C;Accession: T21433
R;Barlow, K.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19421
A;Accession: T21433
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-405 <WIL>
A;Cross-references: EMBL:Z81515; PIDN:CAB04198.1; GSPDB:GN00020; CESP:F26H11.3a
A;Experimental source: clone F26H11
C;Genetics:
A;Gene: CESP:F26H11.3a
A;Map position: 2
A;Introns: 11/1; 42/3; 183/2; 243/2; 287/1
C;Superfamily: bromodomain homology
F;266-321/Domain: bromodomain homology <BRO>

  Query Match             5.3%;  Score 170.5;  DB 2;  Length 405;
  Best Local Similarity   23.4%;  Pred. No. 0.015;
  Matches   61;  Conservative   37;  Mismatches   88;  Indels   75;  Gaps   10;

Qy        133 PASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLRRKREQEPVERSLRGSQNRLRKKRREE 192
              ||:  ||:  :|  : :|     | |   :        :||:| | |:  : :| |:
Db         10 PAAKPKAEVKKEVINPADITLGGDTYDYV-------KEQKPTE-SIATNVSRRRRT---- 57

Qy        193 DSAETGSVQIGSAEQDRPLCKQE---PEASQGPVSQSETDDIENQLEGKATQGN------ 243
                  |  :  :| ||   : :   |:: :   |:     :     | ||  :
Db         58 ------SANLSKSEDDRDKPESQSTAPKSKERRTSEPPASHVAFHTPGSATPHDINLSIE 111

Qy        244 --TEENPREAGKPKPECE----------------------------------VYDPNAL 266
                 | :   :| |   :||                                  | |  ||
Db        112 HCTCQKIFDASKLYIQCELCARWYHGDCVGVAEQTILGLEHWSCEECIEEQERVKDQPAL 171

Qy        267 YCICRQPHNN-RFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDET 325
              ||:|::|::: :| : || |: ||| :||| : |           || || ||       |
Db        172 YCVCQKPYDDTKFYVGCDSCQGWFHPECVGTTRAEA----EQAADYNCPACT-----REA 222

Qy        326 NGSATNEQD-SGCRSVGADGT 345
              |  :   | ||   |    |
Db        223 EGYESEASDVSGSSRVSVQLT 243


RESULT 10
A45988
dentin matrix acidic phosphoprotein AG1 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 25-Oct-1996
C;Accession: A45988
R;George, A.; Sabsay, B.; Simonian, P.A.; Veis, A.
J. Biol. Chem. 268, 12624-12630, 1993
A;Title: Characterization of a novel dentin matrix acidic phosphoprotein. Implications f
A;Reference number: A45988; MUID:93286101; PMID:8509401
A;Accession: A45988
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-489 <GEO>
A;Experimental source: odontoblast-pulp fibroblast, Sprague-Dawley
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:133593, NCBIP:133595)
C;Keywords: phosphoprotein

  Query Match             5.3%;  Score 170.5;  DB 2;  Length 489;
  Best Local Similarity   25.0%;  Pred. No. 0.018;
  Matches   68;  Conservative   50;  Mismatches  101;  Indels   53;  Gaps   11;

Qy          3 DKGH-------LSNEEAPKAIKPTSK---------EFRKTWGFRRTTIAKREGAGDTEAD 46
              |:||       :|:||:    :|||          ||     |||: ::: :  |:  ||

Db        205 DRGHTRMSSADISSEESKGDHEPTSTQDSDDSQDVEPSSRKSFRRSRVSEEDDRGEL-AD 263

Qy         47 PSEQQPQQHNLSLRRSGRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTVTDVE 106
              : :: |  :    || ::  |:|  |:   |   : ::: |   : ||| :       |
Db        264 SNSRETQSVSTEDFRS-KEESRSETQED---TAETQSQEDSPEGQDPSSESS-------E 312

Qy        107 TASEGSVESSSEIRSGPVSDSLGKEHPASSE-----KAKGGEEE------------EDT 148
              | | | ||||| : |  |:| |     :|:      ::  ||:            |:
Db        313 EAGEPSQESSSESQEGVASESRGDNPDNTSQTGDQRDSESSEEDRLNTFSSSESQSTEEQ 372

Qy        149 SDSDS-DGLTLKELQNRLRRKRE---QEPVERSLRGSQNRLRKKRREEDSAETGSVQIGS 204
              ||:| : |:| |       :  :   || ::      ::| :: : ||||    :    |
Db        373 GDSESSNESLSLSEESQESAQDEDSSSQEGLQSQSASRESRSQESQSEEDSRSEENRDSDS 432

Qy        205 AEQDRPLCKQEPEASQGPVSQSETDDIENQLE 236
              :  |   :|   | |  | || |:    :|
Db        433 QDSSR---SKEESNSTGSTSSSEEDNHPKNIE 461


RESULT 11
T13564
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N;Alternate names: hypothetical protein EG:49E4.1
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13564
R;Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A;Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A;Reference number: Z17689
A;Accession: T13564
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5327 <SPA>
A;Cross-references: EMBL:AL031128; PIDN:CAA20006.1
C;Genetics:
A;Cross-references: FlyBase:FBgn0025392
A;Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A;Note: EG:49E4.1
C;Superfamily: Drosophila 576K microtubule-associated protein homolog

  Query Match             5.2%;  Score 167;  DB 2;  Length 5327;
  Best Local Similarity   18.1%;  Pred. No. 0.39;
  Matches  114;  Conservative  103;  Mismatches  270;  Indels  142;  Gaps   23;

Qy          2 DDKGHLSNEEAPKAIK----------PTSKEFRKTWGFRRT-------TIAKREGAGDTE 44
              |:|      |   :::|          | |||  :      :       | ::|    |:
Db       2937 DEKAESRRESVAESVKPESSKDATSAPPSKEHSRPESVLGSLKDEGDKTTSRRVSVADSI 2996

Qy         45 ADPSE-------QQPQQHNLSLRRSGRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEP 97
              |           :|:    ||: :   : | | |    :| : ||   ||: :::| |
Db       2997 KDEKSLLVSQEASRPESEAESLKDAAAPSQETSRPESVTESV-KDGKS--PVASKEASRP 3053

Qy         98 TSSTVTDVETASEGS---VESSSEIRSGPVSDSLGKEHPASSEKA-KGGEEEEDTSDSDS 153
              |      ::| |      ||  : ::| : |   |   || ::| |  ||    | ::
Db       3054 ASVAENAKDSADESKEQRPESLPQSKAGSIKDE--KSPLASKDEAEKSKEESRRESVAEQ 3111

Qy        154 DGLTLKE----------LQNRLRRKREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIG 203
              | ||           :::   : :|: |           |  |     ::  |||:
Db       3112 PPLVSKEVSRPASVAESVKDEAEKSKEESP----------LMSKEASRPASVAGSVK-D 3159

Qy        204 SAEQDRPLCKQEPEASQGP----------VSQSETDDIENQLEGKATQGNTEENP---R 249
              ||: :   ::|  | : |           |::|  |: :   |     :   |::|   :
Db       3160 EAEKSKEESRRESVAEKSPLPSKEASRPASVAESVKDEADKSKEESRRESGAEKSPLASK 3219

Qy        250 EAGKPKPECEVYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEARGRLLERN-- 307
              || :|   |   |                 :: :|             |  : |::
Db       3220 EASRPASVAESIKDEA----------------EKSKE---------ESRRESVAEKSPL 3253

Qy        308 -GEDYICPNCTILQVQDE--------TNGSATNEQDSGCRSVGADGTDCTSIGTVEQKSG 358
               ::   |      |:||        :  |   :    :     :   |:    :||
Db       3254 PSKEASRPTSVAKSVKDEAEKSKEESSRDSVAEKSPLASKEASRPASVAESVQDEAEKSK 3313
```



RESULT 12
T29757
protein UNC-89 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999
C;Accession: T29757
R;Du, Z.; Le, T.T.; Wilson, R.
submitted to the EMBL Data Library, May 1997
A;Description: The sequence of C. elegans cosmid C09D1.
A;Reference number: Z20679
A;Accession: T29757
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6642 <DUZ>
A;Cross-references: EMBL:AF003131; PIDN:AAB54132.1; GSPDB:GN00019; CESP:unc-89
A;Experimental source: strain Bristol N2; clone C09D1
C;Genetics:
A;Gene: CESP:unc-89
A;Map position: 1
A;Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 6
/3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

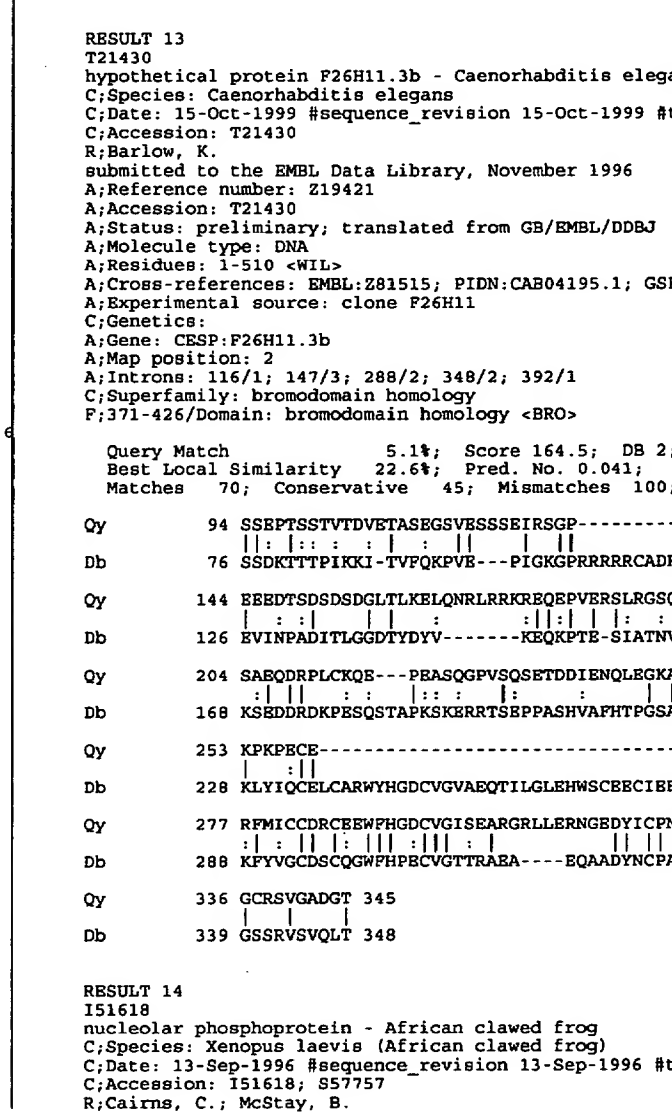
Query Match 5.1%; Score 165; DB 2; Length 6642;
Best Local Similarity 19.9%; Pred. No. 0.65;
Matches 123; Conservative 87; Mismatches 214; Indels 194; Gaps 24;





RESULT 13
T21430
hypothetical protein F26H11.3b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
C;Accession: T21430
R;Barlow, K.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19421
A;Accession: T21430
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-510 <WIL>
A;Cross-references: EMBL:Z81515; PIDN:CAB04195.1; GSPDB:GN00020; CESP:F26H11.3b
A;Experimental source: clone F26H11
C;Genetics:
A;Gene: CESP:F26H11.3b
A;Map position: 2
A;Introns: 116/1; 147/3; 288/2; 348/2; 392/1
C;Superfamily: bromodomain homology
F;371-426/Domain: bromodomain homology <BRO>

Query Match 5.1%; Score 164.5; DB 2; Length 510;
Best Local Similarity 22.6%; Pred. No. 0.041;
Matches 70; Conservative 45; Mismatches 100; Indels 95; Gaps 14;



RESULT 14
I51618
nucleolar phosphoprotein - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Sep-1999
C;Accession: I51618; S57757
R;Cairns, C.; McStay, B.

J. Cell Sci. 108, 3339-3347, 1995
A;Title: Identification and cDNA cloning of a Xenopus nucleolar phosphoprotein, xNopp180
A;Reference number: I51618; MUID:96019267; PMID:7593294
A;Accession: I51618
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-990 <CAI>
A;Cross-references: EMBL:X88927; NID:g895920; PIDN:CAA61368.1; PID:g895921
C;Genetics:
A;Gene: xNopp180
C;Superfamily: nucleolus-cytoplasm shuttle phosphoprotein
C;Keywords: phosphoprotein

Query Match 5.1%; Score 163.5; DB 2; Length 990;
Best Local Similarity 18.8%; Pred. No. 0.098;
Matches 128; Conservative 78; Mismatches 248; Indels 227; Gaps 24;

```
Qy          3 DKGHLSNEE--------------APKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPS 48
              |    |:||               |  | || :|               | | : |:|:| |
Db        297 DSSDSSDEEEQPAKKAKIVPAKAAASAPKPLAK--------------KAETSTDSESDSS 342

Qy         49 EQQPQQHNLSLRRSGRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTVTDVETA 108
               :  :: :: |       ||:     :  :|      ||: |      ||  : |:  :  |
Db        343 SEDEKKSSVKLGVKA-APKKAPAAPDAKSTPVAAAKKSAPAKKASSSSDSDSSSNEETTT 401

Qy        109 SEGSVESSSEIRSGPVS--DSLGKEHPASSEKAKGG----------EEEEDTSDSDSD-G 155
              :  : ::  : | |   : || | |   || |               |:||| ||
Db        402 KPAAKTTPAKSAATPTSKTPTNGKATPTSKTPAKPGTPKTSTAKKDSSSSDSSSDSSSDEE 461

Qy        156 LTLKELQNRLRRKREQEPVERSLRGSQNRLRKKR-----------REEDSAETGSVQIG 203
              | |       |    | ::   |:     |:              ::||: : |
Db        462 TTTKPAAKTTPAKSAATPTSKTPTNSKATPTSKKTPAKPGTPKTSAAKKDSSSSDSSDSS 521

Qy        204 SAEQD-----------------------------RPLCKQEP--------EASQGPVSQ 225
              | |:                            :|  |  |         |:|    |
Db        522 SDEKKTPAKRAAKTTPAKPAAKTTPAKPAAKTTPAKPAAKSTPGKQVPTKKESSSSDSSD 581

Qy        226 SETDDIENQLEG---KATQGNTEENPREAGKPKPECEVYDPNALYCICRQPHNNRFMICC 282
              | ::| :       | | |      |  | ||| |  :    :
Db        582 SSSEDEKKSSAKPAVKTTPGKATSKPVVASKPVPAKKASSSSD----------------S 625

Qy        283 DRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTIL---QVQDETNGSATNEQDSGCRS 339
              |  ||       :::             : |     |   : :  :: |: :: :   :
Db        626 DSSEEETTKTTKPLTK------------LSPAVKTLPPKKAESSSDSSSDSDSEKKTKP 672

Qy        340 VGADGTDCTSIGTV--------------EQKSGEDQGIKGRIEKAANPSGKKKLKIFQPV 385
                    | : |                  | |::|      |:  |:||       |
Db        673 AKPPAKSATPVNTKAPAQNKASKASCSDSDSSSEEEG------KSKQPTGK------SPA 720

Qy        386 VEAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAAATMRFLSSGKEQKTKPKEKVKTKPE 445
              :|   ||   |   :  :| |    |:                ||| ::| |||:    |
Db        721 AKATAPPK-KNPVAVNKDKPSSSSSSD-------------SSGDDEKQKPKQAAAAKDV 765

Qy        446 KFS------LPKCSVQVGIKISSVHKRLASEKREN------PV--KKVMLASRSETSG-- 489
              |        || :     : ||  : ::   |: |      ||    | : |:: |:|
Db        766 KQGAKAAKPTPKKAASSSSEDSSSDEDVSKAKKTNTAVSKSPVTTPKAVPAAKKESSSES 825

Qy        490 ----------------------KEAACESSTPSWASDH------------------NY 507
                                     | || | |   | :|:
Db        826 SDSEDEKQGGKNTSTTKIANSTPKAAAAECSEESSSSEDEGKANGTSGKRKRESTGNAEC 885

Qy        508 NAVKPEKPEKPTALSPTLLSK 528
              || ||  :|  | ||    |
Db        886 EAVTPEN-KKLKAKSPNTFPK 905
```

RESULT 15
T18837
hypothetical protein C01G6.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T18837
R;Berks, M.

submitted to the EMBL Data Library, August 1994
A;Reference number: Z19029
A;Accession: T18837
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-952 <WIL>
A;Cross-references: EMBL:Z35595; PIDN:CAA84636.1; GSPDB:GN00020; CESP:C01G6.5
A;Experimental source: clone C01G6
C;Genetics:
A;Gene: CESP:C01G6.5
A;Map position: 2
A;Introns: 34/2; 61/2; 101/3; 195/1; 218/3; 845/3; 886/3

Query Match 5.0%; Score 161.5; DB 2; Length 952;
Best Local Similarity 23.9%; Pred. No. 0.12;
Matches 60; Conservative 41; Mismatches 111; Indels 39; Gaps 9;

```
Qy         90 SLEDSSEPTSSTVTDVETASEGSVESSSSEI----------RSGPVSDSLGKEHPASSEK 138
              |: :  | :|   || : | : | : |: :             | |   |  |:   || |
Db        678 SVAEEEETSSKENTDQKRAVDDSSDESARLVKQIKALSATPSSTPAESSKRKQKDTSSRK 737

Qy        139 AKGGEEEEDTSDSDSDGLTLKELQNRLRRKREQEPVERSLRGSQNRLRKKRREEDSAETG 198
              |  :|    |||: || :  :   |  : |    : :|    :   :|:: || ||
Db        738 MKQLDESSADSDSEDDGRSGDKSTKRRDKARRSTRIGKSTPAKKVG-KKEKVVEDEDETD 796

Qy        199 SVQIGSAEQDRPLCK------QEPEASQGPVSQSET--DDIENQLEGKATQGNTEENPRE 250
              ||    |:::   :       :  |||  : : :|   |:| :   |  :|  :  | |
Db        797 DVQ----EEEKVTPRGGRRKNMKRTASQSAIKERKTKDKDVEPEEAPKKKRGRKKATPTE 852

Qy        251 AGKP------KPECEVYDPNALYCICR--QPHNNRFMICCDRCEEWFHGDCVGISEARGR 302
              :|      |  | |      :||   :   |  : |  | :|||  || :
Db        853 EEEPPKTEPSKERCGVAKG---HCISAKYEKRKNLQWVSCSICNQWFHVWCVRLD----N 905

Qy        303 LLERNGEDYIC 313
              :  |  | ::|
Db        906 VCYREDETFLC 916
```

Search completed: April 20, 2004, 21:20:24
Job time : 24 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 20, 2004, 20:58:54 ; Search time 17 Seconds
(without alignments)
1880.651 Million cell updates/sec

Title: US-09-787-016A-4
Perfect score: 3228
Sequence: 1 MDDKGHLSNEEAPKAIKPTS.....LKSCVGLMLEAISYFSFRPW 614

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3218	99.7	614	1	DAT1_MOUSE	Q8c9b9 mus musculu
2	2052.5	63.6	1225	1	DAT1_HUMAN	Q9btc0 homo sapien
3	239.5	7.4	2039	1	PHF3_HUMAN	Q92576 homo sapien
4	197.5	6.1	1805	1	NEST_RAT	P21263 rattus norv
5	191.5	5.9	2492	1	ATRX_PONPY	Q7yqm3 pongo pygma
6	185	5.7	1359	1	ATRX_CAEEL	Q9u7e0 caenorhabdi
7	184.5	5.7	2492	1	ATRX_HUMAN	P46100 homo sapien
8	183.5	5.7	2492	1	ATRX_PANTR	Q7yqm4 pan troglod
9	169	5.2	700	1	TRDN_CANFA	P82179 canis famil
10	169	5.2	1227	1	LAP4_HUMAN	P51826 homo sapien
11	167.5	5.2	660	1	CGBP_MOUSE	Q9cww7 mus musculu
12	166.5	5.2	656	1	CGBP_HUMAN	Q9p0u4 homo sapien
13	165.5	5.1	489	1	DMP1_RAT	P98193 rattus norv
14	165.5	5.1	1781	1	AK12_HUMAN	Q02952 homo sapien
15	165	5.1	6632	1	UN89_CAEEL	O01761 caenorhabdi
16	163	5.0	503	1	DMP1_MOUSE	O55188 mus musculu
17	161.5	5.0	952	1	YKI5_CAEEL	P46012 caenorhabdi
18	160	5.0	650	1	SPL1_MOUSE	P70663 mus musculu
19	159.5	4.9	705	1	TRDN_RABIT	Q28820 oryctolagus
20	159	4.9	1096	1	PHF2_MOUSE	Q9wtu0 mus musculu
21	158.5	4.9	751	1	SET3_YEAST	P36124 saccharomyc
22	158	4.9	5147	1	PCLO_HUMAN	Q9y6v0 homo sapien
23	157	4.9	1101	1	PHF2_HUMAN	O75151 homo sapien
24	155.5	4.8	2476	1	ATRX_MOUSE	Q61687 mus musculu
25	155	4.8	728	1	TRDN_HUMAN	Q13061 homo sapien
26	155	4.8	759	1	EPLI_HUMAN	Q9uhb6 homo sapien
27	154.5	4.8	3726	1	ABP1_MOUSE	Q61329 mus musculu
28	154	4.8	1010	1	HC58_MOUSE	Q8blg0 mus musculu
29	153.5	4.8	633	1	MLH_TETTH	P40631 tetrahymena
30	153.5	4.8	1849	1	IGA4_HAEIN	P45386 haemophilus
31	153	4.7	1647	1	PQE1_CAEEL	Q10124 caenorhabdi
32	152	4.7	668	1	SG1_PIG	Q9glg4 sus scrofa
33	150	4.6	510	1	DMP1_BOVIN	Q95120 bos taurus
34	149.5	4.6	835	1	TF1B_HUMAN	Q13263 homo sapien
35	149.5	4.6	2468	1	MAPB_HUMAN	P46821 homo sapien
36	149	4.6	406	1	NSB1_MOUSE	Q9j135 mus musculu
37	148.5	4.6	704	1	NP14_RAT	P41777 rattus norv
38	148	4.6	657	1	MY16_MOUSE	P17564 mus musculu
39	147	4.6	500	1	GAR2_SCHPO	P41891 schizosacch
40	147	4.6	513	1	DMP1_HUMAN	Q13316 homo sapien
41	147	4.6	955	1	T150_HUMAN	Q9y2w1 homo sapien
42	147	4.6	1267	1	HMT1_HUMAN	Q9h9b1 homo sapien
43	147	4.6	1453	1	NKCR_MOUSE	P30415 mus musculu
44	146.5	4.5	644	1	NFM_RABIT	P54938 oryctolagus
45	146	4.5	282	1	YHP0_YEAST	P38806 saccharomyc

ALIGNMENTS

RESULT 1
DAT1_MOUSE
ID DAT1_MOUSE STANDARD; PRT; 614 AA.
AC Q8C9B9; Q8BRG2; Q9WV00;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Death associated transcription factor 1 (Death inducer-obliterator-1)
DE (DIO-1).
GN DATF1 OR DIO1 OR DIDO1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Pre-B cell;
RX MEDLINE=99324176; PubMed=10393935;
RA Garcia-Domingo D., Leonardo E., Grandien A., Martinez P., Albar J.P.,
RA Izpisua-Belmonte J.-C., Martinez-A C.;
RT "DIO-1 is a novel gene involved in onset of apoptosis in vitro, whose
RT misexpression disrupts limb development.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:7992-7997(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo, and Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).

```
CC   -!- FUNCTION: Putative transcription factor, weakly pro-apoptotic when
CC       overexpressed.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic; translocates to the nucleus
CC       after pro-apoptotic stimuli.
CC   -!- TISSUE SPECIFICITY: Ubiquitous. Expressed at intermediate levels.
CC   -!- INDUCTION: Upregulated during apoptosis.
CC   -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ238332; CAB48401.1; -.
DR   EMBL; AK042474; BAC31270.1; -.
DR   EMBL; AK044919; BAC32141.1; -.
DR   MGD; MGI:1344352; Dido1.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0006915; P:apoptosis; IDA.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF00628; PHD; 1.
DR   SMART; SM00249; PHD; 1.
DR   PROSITE; PS01359; ZF_PHD_1; 1.
DR   PROSITE; PS50016; ZF_PHD_2; 1.
KW   Apoptosis; Nuclear protein; Zinc-finger.
FT   DOMAIN      162    170       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      182    190       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   ZN_FING     265    319       PHD-TYPE.
FT   CONFLICT     45     45       V -> A (IN REF. 1).
FT   CONFLICT    331    331       D -> N (IN REF. 1).
FT   CONFLICT    353    353       I -> V (IN REF. 1).
SQ   SEQUENCE   614 AA;  67409 MW;  15ACC5A4BAEF3539 CRC64;

  Query Match             99.7%;  Score 3218;  DB 1;  Length 614;
  Best Local Similarity   99.5%;  Pred. No. 2.4e-165;
  Matches  611;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MDDKGHLSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPSEQQPQQHNLSLR 60
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db          1 MDDKGHLSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEVDPSEQQPQQHNLSLR 60

Qy         61 RSGRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSEIR 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RSGRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSEIR 120

Qy        121 SGPVSDSLGKEHPASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLRRKREQEPVERSLRG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SGPVSDSLGKEHPASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLRRKREQEPVERSLRG 180

Qy        181 SQNRLRKKRREEDSAETGSVQIGSAEQDRPLCKQEPEASQGPVSQSETDDIENQLEGKAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SQNRLRKKRREEDSAETGSVQIGSAEQDRPLCKQEPEASQGPVSQSETDDIENQLEGKAT 240

Qy        241 QGNTEENPREAGKPKPECEVYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEAR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QGNTEENPREAGKPKPECEVYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEAR 300

Qy        301 GRLLERNGEDYICPNCTILQVQDETNGSATNEQDSGCRSVGADGTDCTSIGTVEQKSGED 360
              |||||||||||||||||||||||:||||||||||||||||||:|||||||
Db        301 GRLLERNGEDYICPNCTILQVQDETNGSATDEQDSGCRSVGADGTDCTSIGTIEQKSGED 360

Qy        361 QGIKGRIEKAANPSGKKKLKIFQPVVEAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 QGIKGRIEKAANPSGKKKLKIFQPVVEAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAA 420

Qy        421 ATMRPLSSGKEQKTKPKEKVKTKPEKPSLPKCSVQVGIKISSVHKRLASEKRENPVKKVM 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ATMRPLSSGKEQKTKPKEKVKTKPEKPSLPKCSVQVGIKISSVHKRLASEKRENPVKKVM 480

Qy        481 LASRSETSGKEAACESSTPSWASDHNYNAVKPEKPEKPTALSPTLLSKCTYHPKAGFPGP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LASRSETSGKEAACESSTPSWASDHNYNAVKPEKPEKPTALSPTLLSKCTYHPKAGFPGP 540

Qy        541 SHHLGGCLGLSRTRVLGVLVLIVASSSLPARSRYQDASGPQVFLPSLWSLSGWFLKSCVG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 SHHLGGCLGLSRTRVLGVLVLIVASSSLPARSRYQDASGPQVFLPSLWSLSGWFLKSCVG 600

Qy        601 LMLEAISYFSFRPW 614
              ||||||||||||||
Db        601 LMLEAISYFSFRPW 614


RESULT 2
DAT1_HUMAN
ID   DAT1_HUMAN     STANDARD;      PRT;  1225 AA.
AC   Q9BTC0; O15043; Q9BQW0; Q9BW03; Q9H4G6; Q9H4G7; Q9NTU8; Q9NUM8;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Death associated transcription factor 1.
GN   DATF1 OR KIAA0333.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 3).
RC   TISSUE=Placenta;
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA   Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA   Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA   Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA   Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA   Ninomiya K., Iwayanagi T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21638749; PubMed=11780052;
RA   Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA   Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA   Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA   Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA   Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA   Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA   Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA   Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA   Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA   Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA   Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA   Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA   Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA   Lehvaeslaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA   Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,
RA   Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA   Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA   Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA   Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA   Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA   Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA   Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA   Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA   Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA   Rogers J.;
RT   "The DNA sequence and comparative analysis of human chromosome 20.";
RL   Nature 414:865-871(2001).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORMS 2 AND 3).
RC   TISSUE=Colon, Kidney, and Muscle;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
```

```
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SEQUENCE OF 235-1225 FROM N.A. (ISOFORM 1).
RC   TISSUE=Brain;
RX   MEDLINE=97349984; PubMed=9205841;
RA   Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. VII.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 4:141-150(1997).
CC   -!- FUNCTION: Putative transcription factor, weakly pro-apoptotic when
CC       overexpressed (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic; translocates to the nucleus
CC       after pro-apoptotic stimuli (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=1;
CC         IsoId=Q9BTC0-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q9BTC0-2; Sequence=VSP_007208, VSP_007209, VSP_007210;
CC         Note= No experimental confirmation available;
CC       Name=3;
CC         IsoId=Q9BTC0-3; Sequence=VSP_007208, VSP_007211, VSP_007212;
CC         Note= No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: Ubiquitous.
CC   -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC   -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous
CC       gene model prediction.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AK002127; BAA92094.1; -.
DR   EMBL; AL035669; CAC12755.1; ALT_SEQ.
DR   EMBL; AL035669; CAC12756.1; -.
DR   EMBL; AL035669; CAC28883.1; -.
DR   EMBL; AL117379; CAB90275.1; -.
DR   EMBL; BC000770; AAH00770.1; -.
DR   EMBL; BC004237; AAH04237.1; -.
DR   EMBL; BC014489; AAH14489.1; -.
DR   EMBL; AB002331; BAA20791.1; -.
DR   Genew; HGNC:2680; DATF1.
DR   MIM; 604140; -.
DR   InterPro; IPR003618; TFS2_centre.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF00628; PHD; 1.
DR   SMART; SM00249; PHD; 1.
DR   SMART; SM00510; TFS2M; 1.
DR   PROSITE; PS01359; ZF_PHD_1; 1.
DR   PROSITE; PS50016; ZF_PHD_2; 1.
KW   Apoptosis; Nuclear protein; Zinc-finger; Alternative splicing.
FT   DOMAIN      165    173       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      185    193       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   ZN_FING     268    322       PHD-TYPE.
FT   VARSPLIC    388    423       Missing (in isoform 2 and isoform 3).
FT                                /FTId=VSP_007208.
FT   VARSPLIC    566    580       STKEDRRSEEKAAAT -> CSGKYLYSLHPSLIA (in
FT                                isoform 2).
FT                                /FTId=VSP_007209.
FT   VARSPLIC    581   1225       Missing (in isoform 2).
FT                                /FTId=VSP_007210.
FT   VARSPLIC    566    598       STKEDRRSEEKAAATAASKKTAPPGSTVGKQPA -> CMYH
FT                                LGVGLLDPSRSPWIAIPWACPGLGVAALC (in isoform
FT                                3).
FT                                /FTId=VSP_007211.
FT   VARSPLIC    599   1225       Missing (in isoform 3).
FT                                /FTId=VSP_007212.
FT   CONFLICT    252    252       G -> E (IN REF. 1).
FT   CONFLICT    580    580       T -> M (IN REF. 2).
FT   CONFLICT    592    592       T -> A (IN REF. 2).
SQ   SEQUENCE   1225 AA;  133023 MW;  595071E0D3CC74A2 CRC64;

  Query Match           63.6%;  Score 2052.5;  DB 1;  Length 1225;
  Best Local Similarity  72.2%;  Pred. No. 1.1e-102;
  Matches  412;  Conservative   32;  Mismatches   82;  Indels   45;  Gaps    7;

Qy          1 MDDKGHLSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPSEQQP--QQHNLS 58
              |||||  |||||||||||||||||||||||||||||||||||||||||| |||| |  |  ||  ||
Db          1 MDDKGDPSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDAEADPLEPPPPQQQLGLS 60

Qy         59 LRRSGRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSE 118
              |||||||||||||||:|||  ||||::::|||||| ||||   || ||||||||||:||
Db         61 LRRSGRQPKRTERVEQFLTIARRRGRRSMPVSLEDSGEPTSCPATDAETASEGSVESASE 120

Qy        119 IRSGPVSDSLG-KEHPASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLRRKREQEPVERS 177
              |||| | |   || |||||| |||:: :||||||||||||||||||||||||||||| ||
Db        121 TRSGPQSASTAVKERPASSEKVKGGDDHDDTSDSDSDGLTLKELQNRLRRKREQEPTERP 180

Qy        178 LRGSQNRLRKKRREEDSAET-GSVQIGSAEQDRPLCKQEPEASQGPVSQSETDDIENQLE 236
              |:| |:|||||||||  ||| ||    : |   | |||||  || |||:  || |::||
Db        181 LKGIQSRLRKKRREEGPAETVGSEASDTVEGVLP-SKQEPENDQGVVSQAGKDDRESKLE 239

Qy        237 GKATQGNTEENPREAGKPKPECEVYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGI 296
              ||| |   :| | : |:|||||| ||||||||||||||||||||||||||||||||||||
Db        240 GKAAQDIKDEEPGDLGRPKPECEGYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGI 299

Qy        297 SEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQDSGCRSVGADGTDCTSIGTVEQK 356
              |||||||||||||||||||||||||||||:    ::|::  |   ||||||||||:|||
Db        300 SEARGRLLERNGEDYICPNCTILQVQDETHSETADQQEAKWRPGDADGTDCTSIGTIEQK 359

Qy        357 SGEDQGIKGRIEKAANPSGKKKLKIFQP---------------------------- 384
              | |||||||||||||||||||||||||
Db        360 SSEDQGIKGRIEKAANPSGKKKLKIFQPGPGPVPTQLPVLWQVLEIAVSRSISAFTLLHC 419

Qy        385 ----VVEAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAAATMRFLSSGKEQKTKPKEKV 440
                  |:||||| |||||||  |||||||||||||||||||||:|||||||| |||||:
Db        420 ISCKVIEAPGASKCIGPGCCHVAQPDSVYCSNDCILKHAAATMKFLSSGKEQKPKPKEKM 479

Qy        441 KTKPEKFSLPKCSVQVGIKISSVHKRLASEKRENPVKK-VMLASRSETSGKEAACESSTP 499
              | |||| |||||  | |||||||||| | ||:|  ||| |:: :|||  ||||||||||
Db        480 KMKPEKPSLPKCGAQAGIKISSVHKRPAPEKKETTVKKAVVVPARSEALGKEAACESSTP 539

Qy        500 SWASDHNYNAVKPEKPEKPTALSPTLLSKCT 530
              |||||||||||||||    |   ||:|| | |
Db        540 SWASDHNYNAVKPEKTAAP---SPSLLYKST 567


RESULT 3
PHF3_HUMAN
ID   PHF3_HUMAN     STANDARD;      PRT;  2039 AA.
AC   Q92576; Q9NQ16; Q9UI45;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
```

```
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   PHD finger protein 3.
GN   PHF3 OR KIAA0244.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX   MEDLINE=21845875; PubMed=11856869;
RA   Fischer U., Struss A.-K., Hemmer D., Michel A., Henn W.,
RA   Steudel W.-I., Meese E.;
RT   "PHF3 expression is frequently reduced in glioma.";
RL   Cytogenet. Cell Genet. 94:131-136(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=97191544; PubMed=9039502;
RA   Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayasi Y.,
RA   Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT   "Prediction of the coding sequences of unidentified human genes. VI.
RT   The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT   analysis of cDNA clones from cell line KG-1 and brain.";
RL   DNA Res. 3:321-329(1996).
RN   [3]
RP   SEQUENCE OF 83-2039 FROM N.A.
RA   Matthews L.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- TISSUE SPECIFICITY: Ubiquitous. Expression is significantly
CC       reduced or lost in glioblastomas, glioblastoma cell lines,
CC       anaplastic astrocytomas, and astrocytomas.
CC   -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF091622; AAF21292.1; -.
DR   EMBL; D87685; BAA13438.2; ALT_INIT.
DR   EMBL; AL050329; CAB99358.1; -.
DR   Genew; HGNC:8921; PHF3.
DR   MIM; 607789; -.
DR   InterPro; IPR003618; TFS2_centre.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF00628; PHD; 1.
DR   SMART; SM00249; PHD; 1.
DR   SMART; SM00510; TFS2M; 1.
DR   PROSITE; PS01359; ZF_PHD_1; 1.
DR   PROSITE; PS50016; ZF_PHD_2; 1.
KW   Zinc-finger.
FT   ZN_FING     717    772       PHD-TYPE.
FT   DOMAIN      929   1029       TFS2M.
FT   DOMAIN     1797   1865       PRO-RICH.
SQ   SEQUENCE   2039 AA;  229480 MW;  FDAFF00576005E9B CRC64;

  Query Match             7.4%;  Score 239.5;  DB 1;  Length 2039;
  Best Local Similarity   20.5%;  Pred. No. 1.9e-05;
  Matches  101;  Conservative   75;  Mismatches  163;  Indels  153;  Gaps   17;

Qy         90 SLEDSSEPTSSTVTDVETASEGSVESSSEIRS--GPVSDSLGK--------EHPASSEKA 139
Db        466 NLQDDRNSQSSSVSYLESKSVKSKHTKPVIHSKQNMTTDAPKKIVAAKYEVIHSKTKVNV 525

Qy        140 KGGEEEEDTSDSDSDGLTLKELQN-----RLRRKR-EQEPV-------ERSLRGSQNRLR 186
Db        526 KSVKRNTDVPESQ---------QNFHRPVKVRKKQIDKEPKIQSCNSGVKSVKNQAHSVL 576

Qy        187 KKRREEDSAETGSVQI--------GSAEQDRPLCKQEPE--ASQGPVSQSE--------- 227
Db        577 KKTLQDQTL----VQIFKPLTHSLSDKSHAHPGCLKEPHHPAQTGHVSHSSQKQCHKPQQ 632

Qy        228 -------TDDIENQLEGKATQGNTEENPREAGKPKPECE---------------------- 259
Db        633 QAPAMKTNSHVKEELEHPGVEHFKEEDKLKLKKPEKNLQPRQRRSSKSFSLDEPPLFIPD 692

Qy        260 --------------------VYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEA 299
Db        693 NIATIRREGSDHSSSFESKYMWTPSKQCGPCKKPHGNRFMVGCGRCDDWFHGDCVGLSLS 752

Qy        300 RGRLLERNGEDYICPNC-------TILQVQDETNGSATNEQDSGCRSVGADGTDCTSIGT 352
Db        753 QAQQMGEEDKEYVCVKCCAEEDKKTEILDPDTLENQATVEFHSGDKTM-----ECEKLGL 807

Qy        353 VEQKSGE-----DQGIKGRIEKAANPSGKKKLKIFQPVVEAPGAPKCIGPGCSSVAQPDS 407
Db        808 SKHTTNDRTKYIDDTVKHKVKILKRESGEGR--------------------------NS 840

Qy        408 VYCSNDCILKHAAATMRPLSSGKEQKTKPKEKVKTKPEKFSLPKCSVQVGIKISSVHKRL 467
Db        841 SDCRDNEIKKWQLAPLRKMG----QPVLPRRSSEEKSEK--IPKESTTVTCTGEKASKPG 894

Qy        468 ASEKRENPVKKV 479
Db        895 THEKQEMKKKKV 906


RESULT 4
NEST_RAT
ID   NEST_RAT       STANDARD;      PRT;  1805 AA.
AC   P21263;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Nestin.
GN   NES.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90150286; PubMed=1689217;
RA   Lendahl U., Zimmerman L.B., McKay R.D.G.;
RT   "CNS stem cells express a new class of intermediate filament
RT   protein.";
RL   Cell 60:585-595(1990).
CC   -!- TISSUE SPECIFICITY: CNS STEM CELLS.
CC   -!- DEVELOPMENTAL STAGE: UPON TERMINAL NEURAL DIFFERENTIATION, NESTIN
CC       IS DOWN-REGULATED AND REPLACED BY NEUROFILAMENTS.
CC   -!- SIMILARITY: Belongs to the intermediate filament family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M34384; AAA41685.1; -.
DR   PIR; A34736; A34736.
DR   InterPro; IPR001664; IF.
DR   Pfam; PF00038; filament; 1.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Neurone.
FT   DOMAIN        1      7       HEAD.
FT   DOMAIN        8    314       ROD.
FT   DOMAIN      315   1805       TAIL.
FT   DOMAIN        8     43       COIL 1A.
FT   DOMAIN       44     55       LINKER 1.
FT   DOMAIN       56    151       COIL 1B.
FT   DOMAIN      152    174       LINKER 12.
```

```
FT   DOMAIN       175    193       COIL 2A.
FT   DOMAIN       194    196       LINKER 2.
FT   DOMAIN       197    314       COIL 2B.
SQ   SEQUENCE   1805 AA;  198744 MW;  B40EE14717E0998D CRC64;

  Query Match             6.1%;  Score 197.5;  DB 1;  Length 1805;
  Best Local Similarity   21.2%;  Pred. No. 0.0029;
  Matches  143;  Conservative   99;  Mismatches  223;  Indels  211;  Gaps   35;

Qy          2 DDKG---HLSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPSEQQPQQHNLS 58
              :|:|    |  :||   :|  :   :||          :|   |||     | | |   :
Db        486 EDEGQIWELVEKEADIEVKVENSSAQKT----------QESGLDTEETQDSQGPLQKE-T 534

Qy         59 LRRSGRQPKRTERVEEFLTTVRRRGKKNVPVSLEDS----SEPTSSTVTDVETASEGSVE 114
              |:  | :|  : ::: : |     ||:|   | |         |       :::  | :||
Db        535 LKALGEEPLMSLKIQNYETA----GKENCNSSTEGHLGTLEGPEKEKQIPLKSLEEKNVE 590

Qy        115 SSSEIRSG-PV-SDSLGKEHPASSEK----AKG--------GEEEEDTSDSDSDG----- 155
              |   : :| || |: ||||   : ::     ||        |:| ::     :|
Db        591 SEKTLENGVPVLSELLGKEDTRTEDQELMSPKGTLKRFSSLGKESQEVVRPSKEGNLESW 650

Qy        156 ----------------------LTLKELQN-------------RLRRKREQEP------ 173
                                   |  || |:              |  :|  |||
Db        651 TAFKEESQHPLGFPGAEDQMLERLVEKEDQSFPRSPEEEDQEACRPLQKENQEPLGYEEA 710

Qy        174 ----VERSL-RGSQNRLRKKRREEDSAETGSVQIGSAEQDRPLCKQEPEASQGPVSQSET 228
                  :|| : : ||  ||    ||:  | |           | | |:    :| |:    |
Db        711 EGQILERLIEKESQESLRSP--EEEDQEAG----------RSLQKE----NQEPLGYEEA 754

Qy        229 DD--IENQLEGKATQG--NTEENPREAGKP-----KPECEVYDPNALYCICRQPHNNRFM 279
              :|  :|  :| :: :   : ||| |  |||     :   : |    :  : | | :
Db        755 EDQMLERLIEKESQESLKSPEENQR-IGKPLERENQKSLRYLEENQETFVPLESRNQRPL 813

Qy        280 ICCDRCEE----------------------------WFHGDCVGIS-------EARGRL 303
                :  ||                           ||:  |       :: | |
Db        814 RSLEVEEEEQRIVKPLEKVSQDSLGSLAEENVQPLRYLEEDDCINKSLLEDKTHKSLGSL 873

Qy        304 LERNGEDYICPNCTILQV----------------QDETNGSATNEQDSGCRSVGADGTD 346
              :|||:  | |  :  ||                ::    | ::|::        :| :
Db        874 EDRNGDSIIIPQESETQVSLRPPEEEDQRIVNHLEKESQEFSRSSEEEEQVMERSLEGEN 933

Qy        347 CTSIGTVEQKSGEDQGIKGRIEKAANPSGK----KKLKIFQPVVEAPGAPKCIGPGCSSV 402
              |: :||:   |||  :: ::|| :  |||    :  : | |:       |         |:
Db        934 HESLSSVEK---EDQMVESQLEKESQDSGKSLEDESQETFGPL------EKENAESLRSL 984

Qy        403 AQPDSVYCSNDCILKHAAATMRFLSSGKEQ-KTKPKEKVKTKPEKFSLPKCSVQVGIKIS 461
              | |         : :    |:|  : | ||    | |||   ||   |||    :|
Db        985 AGQDQ---EEQKLEQETQQTLR--AVGNEQMAVSPPEKV--DPE---LPK---PLG-NDQ 1030

Qy        462 SVHKRLASEKREN----------PVKKV------MLASRSETSGKEAACESSTPSWASDH 505
              : : |  | :|:           || :       | :  |   :  : ::  | |:::
Db       1031 EIARSLGKENQESLVSLKEKGIETVKSLETEIIEPLETAEEDLERRKSIDTQEPLWSTEV 1090

Qy        506 NYNAVKPEKPEKPTAL 521
                |:| : | | :|
Db       1091 ARETVEPPEDEPPGSL 1106


RESULT 5
ATRX_PONPY
ID   ATRX_PONPY     STANDARD;      PRT;  2492 AA.
AC   Q7YQM3;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcriptional regulator ATRX (X-linked helicase II) (X-linked
DE   nuclear protein) (XNP).
GN   ATRX.
OS   Pongo pygmaeus (Orangutan).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OX   NCBI_TaxID=9600;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22763540; PubMed=12777533;
RA   Kitano T., Schwarz C., Nickel B., Paabo S.;
RT   "Gene diversity patterns at 10 X-chromosomal loci in humans and
RT   chimpanzees.";
RL   Mol. Biol. Evol. 20:1281-1289(2003).
CC   -!- FUNCTION: Could be a global transcriptional regulator. Modifies
CC       gene expression by affecting chromatin. May be involved in brain
CC       development and facial morphogenesis.
CC   -!- SUBUNIT: Probably binds EZH2. Binds annexin V in a calcium and
CC       phosphatidylcholine/phosphatidylserine-dependent manner (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear. Associated with pericentromeric
CC       heterochromatin during interphase and mitosis, probably by
CC       interacting with HP1.
CC   -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC   -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB102643; BAC81112.1; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
KW   DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding;
KW   Zinc-finger.
FT   ZN_FING     220    268       PHD-TYPE.
FT   NP_BIND    1594   1601       ATP (POTENTIAL).
FT   SITE       1719   1722       DEGH BOX.
FT   DOMAIN      745    750       POLY-SER.
FT   DOMAIN     1151   1156       POLY-SER.
FT   DOMAIN     1166   1169       POLY-LYS.
FT   DOMAIN     1202   1206       POLY-SER.
FT   DOMAIN     1259   1266       POLY-ASP.
FT   DOMAIN     1443   1466       POLY-GLU.
FT   DOMAIN     1499   1502       POLY-GLU.
FT   DOMAIN     1929   1939       POLY-LYS.
FT   DOMAIN     1941   1948       POLY-SER.
FT   DOMAIN     2222   2225       POLY-LYS.
FT   DOMAIN     2262   2265       POLY-GLU.
FT   DOMAIN     2420   2425       POLY-GLN.
SQ   SEQUENCE   2492 AA;  282614 MW;  48EC97106D0CEF20 CRC64;

  Query Match             5.9%;  Score 191.5;  DB 1;  Length 2492;
  Best Local Similarity   18.9%;  Pred. No. 0.0086;
  Matches  108;  Conservative  114;  Mismatches  199;  Indels  151;  Gaps   24;

Qy          3 DKGHLSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPSEQQPQQHNLSLRRS 62
              |:   ::|:  |  |  ::| :||  |::  | | |   :: :| :|: |::  :
Db        972 DQSDETSEDDKKQSKKGTEEKKKTSDFKKKVI-KMEQQYESSSSDGTEKLPEREEIC---- 1026

Qy         63 GRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTVTDVETASEG---SVESSSEI 119
                 ||  ::::   |   :: ||     : | :      ::|    | |   | :|| :
Db       1027 -HFPKGIKQIKNGTTDGEKKNKK-----IRDKTSKKKDELSDYAEKSTGKGDSCDSSEDK 1080

Qy        120 RSGPVSDSLGKEHPASSEKAKGGEEEEDTSDSDSDGLTLKE-----LQNRLRR------- 167
              :|   : : |:|    :   |   : :| | ||::  ::||        ||:|
Db       1081 KS--KNGAYGREKKRCTLLGKSSRKRQDCSSSDTEKYSMKEDGCNSSDKRLKRIELRERR 1138

Qy        168 ----KREQEPVERSLRGS----QNRLRKKRREEDSAETGSVQIGSAEQDRPLCKQEPEAS 219
                  ||  : ::     |     :   ||::  |::  :| :   |: |   :   :
Db       1139 NLSSKRNTKEIQSGSSSSDAEESSEDNKKKKQRTSSKKKAVIV--KEKKRNSLRTSTKRK 1196
```

```
Qy        220 QGPVSQSETDDIENQLEGKATQGNTEENPREAGKPKPECEVYDPNALYCICRQPHNNRFM 279
Db       1197 QADITSSSSSDIEDDDQNSIGEGSSDE---QKIKPVTENLVLSSHTGPC---------- 1242

Qy        280 ICCDRCEEWFHGDCVGISEARGRLL---------ERNGEDYICPNCTILQVQ----DETN 326
Db       1243 --------QSSGD-----EALSKSVPVTVDDDDDDNDPENRIAKKMLLEEIKANLSSDED 1289

Qy        327 GSATNEQDSGCRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLKIFQPVV 386
Db       1290 GSSDDEPEEGKKRTGKQN---------EENPGDEEA-KNQVNSESDSDSEESKK------ 1333

Qy        387 EAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAAATMRFLSSGKEQKTKPKE--KVKTKP 444
Db       1334 -----PR----------------YRHRLLRH-KLTVSDGESGEEKKTKPKEHKEVKGRN 1370

Qy        445 EKFSLPKCSVQVGIKISSVHKRLASEKRENPVKKVMLASRSE-----TSGKEAACESSTP 499
Db       1371 RR-----------KVSSEDSE-DSDFQESGVSEEVSESEDEQRPRTRSAKKAELEENQR 1417

Qy        500 SW--------------ASDHNYNAVKPEKPEK 517
Db       1418 SYKQKKKRRRIKVQEDSSSENKSNSEEEEEEK 1449
```

```
RESULT 6
ATRX_CAEEL
ID   ATRX_CAEEL     STANDARD;      PRT;  1359 AA.
AC   Q9U7E0; O02061;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcriptional regulator ATRX homolog (X-linked nuclear protein-1).
GN   XNP-1 OR B0041.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=99365296; PubMed=10433961;
RA   Villard L., Fontes M., Ewbank J.J.;
RT   "Characterization of xnp-1, a Caenorhabditis elegans gene similar to
RT   the human XNP/ATR-X gene.";
RL   Gene 236:13-19(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Fulton R., Wohldmann P.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Could be a global transcriptional regulator. Modifies
CC       gene expression by affecting chromatin (Potential).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF134186; AAD55361.1; -.
DR   EMBL; AF000196; AAC24256.1; -.
DR   PIR; T34036; T34036.
DR   WormPep; B0041.7; CE17314.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00690; DEAH_ATP_HELICASE; FALSE_NEG.
KW   DNA repair; Hydrolase; Helicase; Nuclear protein; ATP-binding;
KW   DNA-binding.
FT   NP_BIND     496    503       ATP (POTENTIAL).
FT   SITE        636    639       DEAH BOX.
FT   DOMAIN       67     70       POLY-ASP.
FT   DOMAIN      266    272       POLY-GLU.
FT   DOMAIN      276    281       POLY-LYS.
FT   DOMAIN      372    375       POLY-LYS.
FT   DOMAIN      603    608       POLY-LYS.
FT   DOMAIN      859    862       POLY-LYS.
FT   CONFLICT    479    479       C -> F (IN REF. 2).
SQ   SEQUENCE   1359 AA;  156191 MW;  EB4342547D4F4E64 CRC64;
```

```
  Query Match             5.7%;  Score 185;  DB 1;  Length 1359;
  Best Local Similarity   21.4%;  Pred. No. 0.0098;
  Matches  146;  Conservative   88;  Mismatches  217;  Indels  232;  Gaps   33;

Qy          8 SNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPSE--QQPQQHNLSLRRSGRQ 65
Db        210 SESEDEKEVKKSKKKSKK--------VVKKESESEDEAPEKKKTEKRKRSKTSSEESSES 261

Qy         66 PKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTVTDVETASE----GSV----ESSS 117
Db        262 EKSDEEEEEKESSPKPKKKK--PLAVKKLSSDEESEESDVEVLPQKKKRGAVTLISDSED 319

Qy        118 EIRSGPVSDSLGKEHPASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLRRKREQ-EPVER 176
Db        320 EKDQKSESEASDVEEKVSKKKAKKQESSESGSDSSEGSITV----NRKSKKKEKPEKKKK 375

Qy        177 SLRGSQNRLRKKRREEDSAETGSVQIGSAEQDRPLCKQEPEASQGPVSQSETDDIENQLE 236
Db        376 GIIMDSSKLQKETIDAERAE--------KERRKRLEKKQKEFN--GIVLEEGEDLTEMLT 425

Qy        237 GKATQGNTEE---NP------REAGKPKPECEVYDPNALYCICR--QPHNNRFMICC--- 282
Db        426 GTSSQRKLKSVVLDPDSSTVDEESKKP---VEVH--NSLVRILKPHQAHGIQFMYDCACE 480

Qy        283 --DRCEEWFHG----DCVGISEARGRLL------------ERNGED-----YICPNCTIL 319
Db        481 SLDRLDTEGSGGILAHCMGL----GKTLQVITFLHTVLMHEKIGEKCKRVLVVVPKNVII 536

Qy        320 -----------QVQDETNGSATNEQD-----------------SGCRSVGADGTDCTSIG 351
Db        537 NWFKEFQKWLVDNDEELDTIDVNELDSYKTIEDRRRALKAWHSSKTPSVMIIGYDLFRIL 596

Qy        352 TVEQ--KSGEDQGIKGRIEKAANPSGKKKLKIFQPVVEAPGAPKCIGPGCSSVAQPDSVY 409
Db        597 TVEDDPKKKKPKNRNRRLEKAKED--------FRKYLQNPG--------------PDMVV 634

Qy        410 CSNDCILKHAAATMRFLSSGKEQKTKPKEKVKTKPEKPSLPKCSVQVGIKISSVHKRLAS 469
Db        635 CD-------------------------EAHKLKNDDSALSKCMVKI------------- 655

Qy        470 EKRENPVKKVMLASRSETSGKEAACESSTPSWASDHNYNAVKPEKPEKPTALSPTLLSKC 529
Db        656 -----------------LTKRRICLTGTP--------------------LQNNLME-- 674

Qy        530 TYHPKAGFPGPSHHLGGCLGLSRTRVLGVLVLIVASSSLPARSRYQDASGPQV-PLPSLW 588
Db        675 -YHCMVNFVKP-----GLLG-TKTEPANRPVNII------NRGRTKDASPLEVSFMKRRC 721

Qy        589 SLSGWFLKSCVG-----LMLEAI 606
Db        722 HVLYDHLKKCVDRKDYRVLTEAI 744
```

```
RESULT 7
ATRX_HUMAN
ID   ATRX_HUMAN     STANDARD;      PRT;  2492 AA.
AC   P46100; P51068; Q15886; Q9H0Z1; Q9NTS3;
DT   01-NOV-1995 (Rel. 32, Created)
```

```
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcriptional regulator ATRX (X-linked helicase II) (X-linked
DE   nuclear protein) (XNP) (Znf-HX).
GN   ATRX OR RAD54L OR XH2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), VARIANTS SER-596;
RP   GLU-740 AND SER-1860, AND VARIANTS ATR-X.
RX   MEDLINE=97123494; PubMed=8968741;
RA   Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J.,
RA   Gibbons R.J.;
RT   "ATRX encodes a novel member of the SNF2 family of proteins: mutations
RT   point to a common mechanism underlying the ATR-X syndrome.";
RL   Hum. Mol. Genet. 5:1899-1907(1996).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).
RX   MEDLINE=97386582; PubMed=9244431;
RA   Villard L., Lossi A.-M., Cardoso C., Proud V., Chiaroni P.,
RA   Colleaux L., Schwartz C., Fontes M.;
RT   "Determination of the genomic structure of the XNP/ATRX gene encoding
RT   a potential zinc finger helicase.";
RL   Genomics 43:149-155(1997).
RN   [3]
RP   SEQUENCE FROM N.A., AND VARIANTS SER-596 AND GLU-740.
RX   MEDLINE=22763540; PubMed=12777533;
RA   Kitano T., Schwarz C., Nickel B., Paabo S.;
RT   "Gene diversity patterns at 10 X-chromosomal loci in humans and
RT   chimpanzees.";
RL   Mol. Biol. Evol. 20:1281-1289(2003).
RN   [4]
RP   SEQUENCE OF 860-2492 FROM N.A.
RX   MEDLINE=95179111; PubMed=7874112;
RA   Stayton C.L., Dabovic B., Gulisano M., Gecz J., Broccoli V.,
RA   Giovanazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E.,
RA   Bianchi M.E., Consalez G.G.;
RT   "Cloning and characterization of a new human Xq13 gene, encoding a
RT   putative helicase.";
RL   Hum. Mol. Genet. 3:1957-1964(1994).
RN   [5]
RP   PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RX   MEDLINE=94214473; PubMed=8162050;
RA   Gecz J., Pollard H., Consalez G., Villard L., Stayton C.L.,
RA   Millasseau P., Khrestchatisky M., Fontes M.;
RT   "Cloning and expression of the murine homologue of a putative human
RT   X-linked nuclear protein gene closely linked to PGK1 in Xq13.3.";
RL   Hum. Mol. Genet. 3:39-44(1994).
RN   [6]
RP   SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.
RX   MEDLINE=95211835; PubMed=7697714;
RA   Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;
RT   "Mutations in a putative global transcriptional regulator cause X-
RT   linked mental retardation with alpha-thalassemia (ATR-X syndrome).";
RL   Cell 80:837-845(1995).
RN   [7]
RP   SEQUENCE OF 1375-2492 FROM N.A.
RA   Pearce A., Chapman J.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   EZH2 BINDING.
RX   MEDLINE=98167853; PubMed=9499421;
RA   Cardoso C., Timsit S., Villard L., Khrestchatisky M., Fontes M.,
RA   Colleaux L.;
RT   "Specific interaction between the XNP/ATR-X gene product and the SET
RT   domain of the human EZH2 protein.";
RL   Hum. Mol. Genet. 7:679-684(1998).
RN   [9]
RP   SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
RP   HETEROCHROMATIN.
RX   MEDLINE=20040663; PubMed=10570185;
RA   McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
RA   Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,
RA   Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
RT   "Localization of a putative transcriptional regulator (ATRX) at
RT   pericentromeric heterochromatin and the short arms of acrocentric
RT   chromosomes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
RN   [10]
RP   DISEASE.
RX   MEDLINE=20213147; PubMed=10751095;
RA   Villard L., Fontes M., Ades L.C., Gecz J.;
RT   "Identification of a mutation in the XNP/ATR-X gene in a family
RT   reported as Smith-Fineman-Myers syndrome.";
RL   Am. J. Med. Genet. 91:83-85(2000).
RN   [11]
RP   DISEASE.
RX   MEDLINE=22804448; PubMed=12858175;
RA   Gibbons R.J., Pellagatti A., Garrick D., Wood W.G., Malik N.,
RA   Ayyub H., Langford C., Boultwood J., Wainscoat J.S., Higgs D.R.;
RT   "Identification of acquired somatic mutations in the gene encoding
RT   chromatin-remodeling factor ATRX in the alpha-thalassemia
RT   myelodysplasia syndrome (ATMDS).";
RL   Nat. Genet. 34:446-449(2003).
RN   [12]
RP   VARIANT ATR-X SER-1713.
RX   MEDLINE=97196774; PubMed=9043863;
RA   Villard L., Lacombe D., Fontes M.;
RT   "A point mutation in the XNP gene, associated with an ATR-X phenotype
RT   without alpha-thalassemia.";
RL   Eur. J. Hum. Genet. 4:316-320(1996).
RN   [13]
RP   VARIANT JM GLN-2131.
RX   MEDLINE=96224392; PubMed=8630485;
RA   Villard L., Gecz J., Mattei J.-F., Fontes M., Saugier-Veber P.,
RA   Munnich A., Lyonnet S.;
RT   "XNP mutation in a large family with Juberg-Marsidi syndrome.";
RL   Nat. Genet. 12:359-360(1996).
RN   [14]
RP   VARIANTS ATR-X.
RX   MEDLINE=97467722; PubMed=9326931;
RA   Gibbons R.J., Bachoo S., Picketts D.J., Aftimos S., Asenbauer B.,
RA   Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K.,
RA   Levin M.L., Masuno M., Neri G., Pierpont M.E., Slaney S.F.,
RA   Higgs D.R.;
RT   "Mutations in transcriptional regulator ATRX establish the functional
RT   significance of a PHD-like domain.";
RL   Nat. Genet. 17:146-148(1997).
RN   [15]
RP   VARIANT ATR-X LEU-246.
RX   MEDLINE=20123062; PubMed=10660327;
RA   Fichera M., Romano C., Castiglia L., Failla P., Ruberto C., Amata S.,
RA   Greco D., Cardoso C., Fontes M., Ragusa A.;
RT   "New mutations in XNP/ATR-X gene: a further contribution to
RT   genotype/phenotype relationship in ATR/X syndrome.";
RL   Hum. Mutat. 12:214-214(1998).
RN   [16]
RP   VARIANT SHS LYS-1742.
RX   MEDLINE=99347960; PubMed=10417298;
RA   Lossi A.-M., Millan J.M., Villard L., Orellana C., Cardoso C.,
RA   Prieto F., Fontes M., Martinez F.;
RT   "Mutation of the XNP/ATR-X gene in a family with severe mental
RT   retardation, spastic paraplegia and skewed pattern of X inactivation:
RT   demonstration that the mutation is involved in the inactivation
RT   bias.";
RL   Am. J. Hum. Genet. 65:558-562(1999).
RN   [17]
RP   VARIANT CWS THR-2050.
RX   MEDLINE=99326061; PubMed=10398237;
RA   Abidi F., Schwartz C.E., Carpenter N.J., Villard L., Fontes M.,
RA   Curtis M.;
RT   "Carpenter-Waziri syndrome results from a mutation in XNP.";
RL   Am. J. Med. Genet. 85:249-251(1999).
RN   [18]
```

```
RP   VARIANTS ATR-X GLU-175; 178-VAL--LYS-198 DEL; SER-190; PRO-219;
RP   LEU-246 AND CYS-249.
RX   MEDLINE=99219535; PubMed=10204841;
RA   Villard L., Bonino M.-C., Abidi F., Ragusa A., Belougne J.,
RA   Lossi A.-M., Seaver L., Bonnefont J.-P., Romano C., Fichera M.,
RA   Lacombe D., Hanauer A., Philip N., Schwartz C.E., Fontes M.;
RT   "Evaluation of a mutation screening strategy for sporadic cases of
RT   ATR-X syndrome.";
RL   J. Med. Genet. 36:183-186(1999).
RN   [19]
RP   VARIANTS ATR-X SER-179; LEU-190; ILE-194; CYS-246; PHE-1552; SER-1645
RP   AND CYS-1847.
RX   MEDLINE=20451413; PubMed=10995512;
RA   Wada T., Kubota T., Fukushima Y., Saitoh S.;
RT   "Molecular genetic study of Japanese patients with X-linked alpha-
RT   thalassemia/mental retardation syndrome (ATR-X).";
RL   Am. J. Med. Genet. 94:242-248(2000).
RN   [20]
RP   VARIANT ATR-X MET-1621.
RX   MEDLINE=22108790; PubMed=12116232;
RA   Yntema H.G., Poppelaars F.A., Derksen E., Oudakker A.R.,
RA   van Roosmalen T., Jacobs A., Obbema H., Brunner H.G., Hamel B.C.J.,
RA   van Bokhoven H.;
RT   "Expanding phenotype of XNP mutations: mild to moderate mental
RT   retardation.";
RL   Am. J. Med. Genet. 110:243-247(2002).
CC   -!- FUNCTION: Could be a global transcriptional regulator. Modifies
CC       gene expression by affecting chromatin. May be involved in brain
CC       development and facial morphogenesis.
CC   -!- SUBUNIT: Probably binds EZH2. Binds annexin V in a calcium and
CC       phosphatidylcholine/phosphatidylserine-dependent manner (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear. Associated with pericentromeric
CC       heterochromatin during interphase and mitosis, probably by
CC       interacting with HP1.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=5;
CC       Name=4;
CC         IsoId=P46100-1; Sequence=Displayed;
CC       Name=1;
CC         IsoId=P46100-2; Sequence=VSP_000575;
CC       Name=2;
CC         IsoId=P46100-3; Sequence=VSP_000574;
CC       Name=3;
CC         IsoId=P46100-4; Sequence=VSP_000576;
CC       Name=5;
CC         IsoId=P46100-5; Sequence=VSP_000574, VSP_000576;
CC   -!- TISSUE SPECIFICITY: Ubiquitous.
CC   -!- DISEASE: Defects in ATRX are the cause of X-linked alpha-
CC       thalassemia/mental retardation syndrome (ATR-X) [MIM:301040]. ATR-
CC       X is an X-linked disorder comprising severe psychomotor

  Query Match             5.7%;  Score 184.5;  DB 1;  Length 2492;
  Best Local Similarity   18.7%;  Pred. No. 0.02;
  Matches  107;  Conservative  113;  Mismatches  201;  Indels  151;  Gaps   24;

Qy          3 DKGHLSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPSEQQPQQHNLSLRRS 62
              |:   ::|:  |  |  ::| :|   |::  | | |   :: :| :|:| |::  :
Db        972 DQSDETSEDDKKQSKKGTEEKKKPSDFKKKVI-KMEQQYESSSDGTEKLPEREEIC---- 1026

Qy         63 GRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTVTDVETASEG---SVESSSEI 119
                 ||  ::::   |   :: ||     : | :      ::|     | |   | :|| :
Db       1027 -HFPKGIKQIKNGTTDGEKKSKK-----IRDKTSKKKDELSDYAEKSTGKGDSCDSSEDK 1080

Qy        120 RSGPVSDSLGKEHPASSEKAKGGEEEEDTSDSDSDGLTLKR-----LQNRLRR------- 167
              :|   : : |:|         |   : :| | ||::  ::||         ||:|
Db       1081 KS--KNGAYGREKKRCKLLGKSSRKRQDCSSSDTEKYSMKEDGCNSSDKRLKRIELRERR 1138

Qy        168 ----KREQEPVERSLRGS----QNRLRKKRREEDSAETGSVQIGSAEQDRPLCKQEPEAS 219
                  ||  : ::      |      :   ||::: |::  :| :   |: |   :   :
Db       1139 NLSSKRNTKEIQSGSSSSDAEESSEDNKKKKQRTSSKKKAVIV--KEKKRNSLRTSTKRK 1196

Qy        220 QGPVSQSETDDIENQLEGKATQGNTEENPREAGKPKPECEVYDPNALYCICRQPHNNRFM 279
```

```
              |  :: | : |||:  :    :|:::|   :  ||  |  |   :  :|
Db       1197 QADITSSSSSDIEDDDQNSIGEGSSDE---QKIKPVTENLVLSSHTGFC---------- 1242

Qy        280 ICCDRCEEWFHGDCVGISEARGRLL---------ERNGEDYICPNCTILQVQ----DETN 326
                        ||      ||  : :          : : |: |      : :::     : :
Db       1243 --------QSSGD-----EALSKSVPVTVDDDDDDNDPENRIAKKMLLEEIKANLSSDED 1289

Qy        327 GSATNEQDSGCRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLKIFQPVV 386
              ||: :| : | :  |             |:  |:::  | ::   ::   ::  |
Db       1290 GSSDDEPEEGKKRTGKQN---------EENPGDEEA-KNQVNSESDSDSEESKK------ 1333

Qy        387 EAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAAATMRFLSSGKEQKTKPKE--KVKTKP 444
                |:                    :|:|   |:    ||:|:||||||  :|| :
Db       1334 -----PR-----------------YRHRLLRH-KLTVSDGESGEEKKTKPKEHKEVKGRN 1370

Qy        445 EKPSLPKCSVQVGIKISSVHKRLASEKRENPVKKVMLASRSE-----TSGKEAACESSTP 499
               :            |:||      |: :|: | : :  |  |       | |:|  | :
Db       1371 RR-----------KVSSEDSE-DSDFQESGVSEEVSESEDEQRPRTRSAKKAELEENQR 1417

Qy        500 SW-------------ASDHNYNAVKPEKPEK 517
              |:              :|  | :  : |: ||
Db       1418 SYKQKKKRRRIKVQEDSSSENKSNSEEEEEEK 1449
```

```
RESULT 8
ATRX_PANTR
ID   ATRX_PANTR     STANDARD;      PRT;  2492 AA.
AC   Q7YQM4;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcriptional regulator ATRX (X-linked helicase II) (X-linked
DE   nuclear protein) (XNP).
GN   ATRX.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22763540; PubMed=12777533;
RA   Kitano T., Schwarz C., Nickel B., Paabo S.;
RT   "Gene diversity patterns at 10 X-chromosomal loci in humans and
RT   chimpanzees.";
RL   Mol. Biol. Evol. 20:1281-1289(2003).
CC   -!- FUNCTION: Could be a global transcriptional regulator. Modifies
CC       gene expression by affecting chromatin. May be involved in brain
CC       development and facial morphogenesis.
CC   -!- SUBUNIT: Probably binds EZH2. Binds annexin V in a calcium and
CC       phosphatidylcholine/phosphatidylserine-dependent manner (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear. Associated with pericentromeric
CC       heterochromatin during interphase and mitosis, probably by
CC       interacting with HP1.
CC   -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC   -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB102642; BAC81111.1; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
```

```
KW   DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding;
KW   Zinc-finger.
FT   ZN_FING      220    268       PHD-TYPE.
FT   NP_BIND     1594   1601       ATP (POTENTIAL).
FT   SITE        1719   1722       DEGH BOX.
FT   DOMAIN       745    750       POLY-SER.
FT   DOMAIN      1151   1156       POLY-SER.
FT   DOMAIN      1166   1169       POLY-LYS.
FT   DOMAIN      1202   1206       POLY-SER.
FT   DOMAIN      1259   1266       POLY-ASP.
FT   DOMAIN      1443   1466       POLY-GLU.
FT   DOMAIN      1499   1502       POLY-GLU.
FT   DOMAIN      1929   1939       POLY-LYS.
FT   DOMAIN      1941   1948       POLY-SER.
FT   DOMAIN      2222   2225       POLY-LYS.
FT   DOMAIN      2262   2265       POLY-GLU.
FT   DOMAIN      2420   2425       POLY-GLN.
SQ   SEQUENCE   2492 AA;  282567 MW;  56978AC7D37400DC CRC64;

  Query Match             5.7%;  Score 183.5;  DB 1;  Length 2492;
  Best Local Similarity  18.7%;  Pred. No. 0.023;
  Matches  107;  Conservative  113;  Mismatches  201;  Indels  151;  Gaps   24;

Qy          3 DKGHLSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPSEQQPQQHNLSLRRS 62
              |:   ::|:  |  |  ::| :|   |::  | | |   :: :| :|: |::  :
Db        972 DQSDETSEDDKKQSKKGTEEKKKPSDFKKKVI-KMEQQYESSSDGTEKLPEREEIC---- 1026

Qy         63 GRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTVTDVETASEG---SVESSSEI 119
               ||  ::::   |   :: ||     : | :       ::|     | |   | :|| :
Db       1027 -HPPKGIKQIKNGTTDGEKKSKK-----IRDKTSKKKDELSDYAEKSTGKGDSCDSSEDK 1080

Qy        120 RSGPVSDSLGKEHPASSEKAKGGEEEEDTSDSDSDGLTLKE-----LQNRLRR------- 167
              :|   : : |:|         |   : :| | ||::  ::||         ||:|
Db       1081 KS--KNGAYGREKKRCKLLGKSSRKRQDCSSSDTEKYSMKEDGCNSSDKRLKRIELRERR 1138

Qy        168 ----KREQEPVERSLRGS----QNRLRKKRREEDSAETGSVQIGSAEQDRPLCKQEPEAS 219
                  ||  : ::     |     :   ||:::  |::  :| :   |: |   :   :
Db       1139 NLSSKRNTKEIQSGSSSSDAEESSEDNKKKQQRTSSKKKAVIV--KEKKRNSLRTSTKRK 1196

Qy        220 QGPVSQSETDDIENQLEGKATQGNTEENPREAGKPKPECEVYDPNALYCICRQPHNNRFM 279
              |  :: | : |||:  :    :|:::|   :  ||  |  |   :  :|
Db       1197 QADITSSSSSDIEDDDQNSIGEGSSDE---QKIKPVTENLVLSSHTGPC----------- 1242

Qy        280 ICCDRCEEWFHGDCVGISEARGRLL---------ERNGEDYICPNCTILQVQ----DETN 326
                        ||      ||  : :          : : |: |      : :::     : :
Db       1243 --------QSSGD-----EALSKSVPVTVDDDDDDNDPENRIAKKMLLEEIKANLSSDED 1289

Qy        327 GSATNEQDSGCRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLKIFQPVV 386
              ||: :| : | :  |             |:  |:::  | ::   ::   ::  |
Db       1290 GSSDDEPEEGKKRTGKQN---------EENPGDEEA-KNQVNSESDSDSEESKK------ 1333

Qy        387 EAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAAATMRPLSSGKEQKTKPKE--KVKTKP 444
                |:                     :|:|   |:    ||:|:|||||  :|| :
Db       1334 -----PR-----------------YRHRLLRH-KLTVSDGESGEEKKTKPKEHKEVKGRN 1370

Qy        445 EKFSLPKCSVQVGIKISSVHKRLASEKRENPVKKVMLASRSE-----TSGKEAACESSTP 499
               :          |:||       |: :|: | : :   |  |       | |:|  | :
Db       1371 RR------------KVSSEDSE-DSDFQESGVSEEVSESEDEQRPRTRSAKKAELEENQR 1417

Qy        500 SW--------------ASDHNYNAVKPEKPEK 517
              |:              :|  | :  : |: ||
Db       1418 SYKQKKKRRRIKVQEDSSSENKSNSEEEEEEK 1449


RESULT 9
TRDN_CANFA
ID   TRDN_CANFA     STANDARD;      PRT;   700 AA.
AC   P82179;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Triadin.
GN   TRDN.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Heart, and Skeletal muscle;
RX   MEDLINE=99428545; PubMed=10497235;
RA   Kobayashi Y.M., Jones L.R.;
RT   "Identification of triadin 1 as the predominant triadin isoform
RT   expressed in mammalian myocardium.";
RL   J. Biol. Chem. 274:28660-28668(1999).
CC   -!- FUNCTION: May be involved in anchoring calsequestrin to the
CC       junctional sarcoplasmic reticulum and allowing its functional
CC       coupling with the ryanodine receptor (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic
CC       reticulum.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=Skeletal;
CC         IsoId=P82179-1; Sequence=Displayed;
CC       Name=Cardiac 1;
CC         IsoId=P82179-2; Sequence=VSP_004001, VSP_004002;
CC       Name=Cardiac 3;
CC         IsoId=P82179-3; Sequence=VSP_004003, VSP_004004;
CC   -!- TISSUE SPECIFICITY: Skeletal and cardiac muscle.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF165916; AAF00222.1; -.
DR   EMBL; AF165915; AAF00221.1; -.
DR   EMBL; AF165917; AAF00223.1; -.
KW   Transmembrane; Sarcoplasmic reticulum; Glycoprotein;
KW   Alternative splicing.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   DOMAIN        1     46       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     47     67       POTENTIAL.
FT   DOMAIN       68    700       LUMENAL (POTENTIAL).
FT   CARBOHYD     74     74       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    616    616       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    257    277       DQYAFCRYMIDMFVHGDLRPG -> GKHSEEVAGGSKRTLG
FT                                KKQIQ (in isoform Cardiac 1).
FT                                /FTId=VSP_004001.
FT   VARSPLIC    278    700       Missing (in isoform Cardiac 1).
FT                                /FTId=VSP_004002.
FT   VARSPLIC    466    466       E -> EPIKGKEVKVPGSLKEKE (in isoform
FT                                Cardiac 3).
FT                                /FTId=VSP_004003.
FT   VARSPLIC    524    579       EEKVVKQVKATEKAAIEKTVKPKPAKKAEHQEKESPTIKTD
FT                                KPKPTSKETPEVTES -> GILQVVPVVLNCLFLVQPQQDE
FT                                ELNVESKVFRMIHVLSHPTSRTSPILVISTTCRT (in
FT                                isoform Cardiac 3).
FT                                /FTId=VSP_004004.
SQ   SEQUENCE   700 AA;  78152 MW;  F033E3AA1BEE0C56 CRC64;

  Query Match             5.2%;  Score 169;  DB 1;  Length 700;
  Best Local Similarity  19.9%;  Pred. No. 0.034;
  Matches  114;  Conservative   84;  Mismatches  190;  Indels  184;  Gaps   24;

Qy          3 DKGHLS----------NEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPSEQQP 52
              ||| :             |:| |  ||  |   |         :| ||      | : ||::
Db        129 DKGEIEEPPLKQKEIHKEKAEKEEKPERKILAK--------VAHREKEKVKEKEKSEKK- 179

Qy         53 QQHNLSLRRSGRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTVTDV------- 105
                 | : :  : :|  ||       : || |    ::  :||:: | :|
Db        180 ATHKEKIEKKEKPETKTMAKEERKAKTEEKIKKEVKGGKQEKVKPTAAKVKEVQKTPPKA 239
```

```
Qy        106 ------ETASEGSVESS---------------SEIRSG------PVSDSLGKEHPASSEK 138
Db        240 KEKEGKETAAVAKHEQKDQYAFCRYMIDMFVHGDLRPGQSPALPPPLPTVQASRPTPASP 299

Qy        139 AKGGEEEEDTSDSDSDGLTLKELQNRLRRKREQEPVERSLRGSQNRLRKKRREEDSAETG 198
Db        300 TLEGKEEEEKKKAE------KKVTSE-TKKKEKEDV------------KKKSDKDTAI-- 338

Qy        199 SVQIGSAEQDRPLCKQEPEASQGP---VSQSETDDIENQLEGKATQGNTEENPREAGKPK 255
Db        339 -----DVEKKEP--GKAPETKQGTIKVVAQAAAKKDEKKEDSKKTKTPVEEHPK--GKKQ 389

Qy        256 PECEVYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPN 315
Db        390 EKKEKY---------------------------------VEPAKSSKKEHSAPSEKQVK 415

Qy        316 CTILQVQDETNGSATNEQDSGCRSVGADGTDCTSIGTVEQK-SGEDQGIKGRIEKAANPS 374
Db        416 AKTERAKEETSAASTKKAVPGKK-------EEKTTKTVEQEIRKEKSGKTSTASKDKEPE 468

Qy        375 GKKKLKIFQPVVEAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAAATMRPLSSGKEQKT 434
Db        469 IKKDEKM-------PKADKEVKP-----KPPQS------------------QVKKEEKS 497

Qy        435 KPKEKVKTKPEKFSLPKCSVQVGIKISSVHKRLASEKRENPVKKVMLASRSETSGKEAAC 494
Db        498 ESQVKKEAKPEQ------------DIAKPEKTVSHGKPEEKVVKQVKAT------EKAAI 539

Qy        495 ESST---PSWASDHNYN---AVKPEKPEKPTA 520
Db        540 EKTVKPKPAKKAEHQEKESPTIKTDKP-KPTS 570
```

```
RESULT 10
LAF4_HUMAN
ID   LAF4_HUMAN     STANDARD;      PRT;  1227 AA.
AC   P51826;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   LAF-4 protein (Lymphoid nuclear protein related to AF4).
GN   LAF4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96141096; PubMed=8555498;
RA   Ma C., Staudt L.M.;
RT   "LAF-4 encodes a lymphoid nuclear protein with transactivation
RT   potential that is homologous to AF-4, the gene fused to MLL in
RT   t(4;11) leukemias.";
RL   Blood 87:734-745(1996).
CC   -!- FUNCTION: Putative transcription activator that may function in
CC       lymphoid development and oncogenesis. Binds, in vitro, to
CC       double-stranded DNA.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: Preferentially expressed in lymphoid tissues,
CC       highest levels being found in the thymus.
CC   -!- SIMILARITY: Belongs to the AF4 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U34360; AAA98763.1; -.
DR   TRANSFAC; T04866; -.
DR   Genew; HGNC:6473; LAF4.
DR   MIM; 601464; -.
DR   GO; GO:0005634; C:nucleus; TAS.
DR   GO; GO:0007275; P:development; TAS.
DR   InterPro; IPR007797; AF-4.
DR   Pfam; PF05110; AF-4; 1.
KW   Nuclear protein; Transcription regulation; Activator; DNA-binding.
FT   DOMAIN      413    419       POLY-SER.
FT   DOMAIN      422    432       POLY-SER.
FT   DOMAIN      440    445       POLY-SER.
FT   DOMAIN      670    679       POLY-SER.
SQ   SEQUENCE   1227 AA;  133734 MW;  634B896FD7E9BBE7 CRC64;
```

```
  Query Match             5.2%;  Score 169;  DB 1;  Length 1227;
  Best Local Similarity   19.1%;  Pred. No. 0.063;
  Matches  144;  Conservative   95;  Mismatches  209;  Indels  306;  Gaps   34;

Qy         13 PKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPSEQQPQQHNLSLRRS---GRQPKRT 69
Db        133 PAAV-PVQQSKRGTMGWQK--------AGHPPSDGQQRATQQGSLRTLLGDGVGRQQPRA 183

Qy         70 ER---VEEFLTTVRR------------RGKKNVPVSL---------------------- 91
Db        184 KQVCNVEVGLQTQERPPAMAAKHSSSGHCVQNFPPSLASKPSLVQQKPTAYVRPMDGQDQ 243

Qy         92 --EDSSEPTSSTVTDVETASEGSVESS--------------------SEIRSG------- 122
Db        244 APDESPKLKSSSETSVHCTSYRGVPASKPEPARAKAKLSKFSIPKQGEESRSGETNSCVE 303

Qy        123 ----------PVS--DSLGKEHP--------------ASSEKAKGGEEEEDTSDSDSDGL 156
Db        304 EIIREMTWLPPLSAIQAPGKVEPTKPPFPNKDSQLVSSGHNNPKKGDAEPESPDNGTSNT 363

Qy        157 TLKELQNRLRRKREQEPVERSLRGSQNRLRKKRREEDSAETGSVQI--GSAEQDRPLCKQ 214
Db        364 SM--LEDDLKLSSDEE----------------ENEQQAAQRTALRALSDSAVVQQPNCRT 405

Qy        215 EPEASQGPVSQSETDDIENQLEGKATQGNTEE-----NPREAGKP----KPECEVYDPNA 265
Db        406 SVPSSKGSSSSSSSGTSSSSDSESSSGSDSETESSSSESEGSKPPHFSSPEAE------ 459

Qy        266 LYCICRQPHNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDET 325
Db        460 ------PASSNKWQL-----DKWL--------------------NKVNPHKPPILIQNES 488

Qy        326 NGSATNEQDSGCRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLKIFQPV 385
Db        489 HGSESNQY---YNPVKEDVQDCGKVPDVCQPSLREKEIKSTCKEEQRPRTANK------- 538

Qy        386 VEAPGA--------PKCIGPGCSSVAQPDSVYCSNDCILKHAAATMRFLSSGKE-----Q 432
Db        539 --APGSKGVKQKSPPAAVAVAVSAAAPPPAVPCAP---AENAPAPAR-RSAGKKPTRRTE 592

Qy        433 KTK------------------------PKEKVKTKP-----------------EKFS- 448
Db        593 RTSAGDGANCHRPEEPAAADALGTSVVVPPEPTKTRPCGNNRASHRKELRSSVTCEKRRT 652

Qy        449 ------LPK----CSVQVGIKISSVHKRLASEKRENPVKKVMLASRSETSG-----KEAA 493
Db        653 RGLSRIVPKSKEPIETESSSSSSSSDSDLESEQEEYPLSKAQTVAASASSGNDQRLKEAA 712

Qy        494 CESSTPSWASDHNYNA-----VKPEKPEKPTALSP------------------------ 523
Db        713 ANGGSGPRAPVGSINARTTSDIAKELEEQFYTLVPFGRNELLSPLKDSDEIRSLWVKIDL 772

Qy        524 TLLSKCTYH-PKAGFPG--------------PSH 542
Db        773 TLLSRIPEHLPQE--PGVLSAPATKDSESAPPSH 804
```

```
RESULT 11
CGBP_MOUSE
ID   CGBP_MOUSE     STANDARD;      PRT;   660 AA.
AC   Q9CWW7;
```

```
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   CpG binding protein (Protein containing PHD finger and CXXC domain 1).
GN   CGBP OR PCCX1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryonic stem cells;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- FUNCTION: Transcriptional activator that exhibits a unique DNA
CC       binding specificity for [AC]CpG[AC] unmethylated CpG motifs (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC   -!- SIMILARITY: Contains 1 CXXC-type zinc finger.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AK010337; BAB26862.1; -.
DR   MGD; MGI:1921572; Cgbp.
DR   GO; GO:0016363; C:nuclear matrix; IDA.
DR   InterPro; IPR002857; Znf_CXXC.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF00628; PHD; 1.
DR   Pfam; PF02008; zf-CXXC; 1.
DR   SMART; SM00249; PHD; 1.
DR   PROSITE; PS01359; ZF_PHD_1; 1.
DR   PROSITE; PS50016; ZF_PHD_2; 1.
KW   Transcription regulation; Activator; DNA-binding; Zinc-finger; Zinc;
KW   Metal-binding; Coiled coil; Nuclear protein.
FT   ZN_FING      28     76       PHD-TYPE.
FT   ZN_FING     165    212       CXXC-TYPE.
FT   DOMAIN      260    321       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      325    363       ARG/LYS-RICH (BASIC).
FT   DOMAIN      426    479       COILED COIL (POTENTIAL).
FT   DOMAIN      156    161       POLY-GLN.
FT   DNA_BIND    166    225       BY SIMILARITY.
SQ   SEQUENCE   660 AA;  76166 MW;  9D92CB6DA82069EA CRC64;

  Query Match             5.2%;  Score 167.5;  DB 1;  Length 660;
  Best Local Similarity   18.4%;  Pred. No. 0.038;
  Matches   76;  Conservative   29;  Mismatches   72;  Indels  237;  Gaps   13;
```

```
Qy      235 LEGKATQGNTEENPREAGKPKPECEVYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCV 294
            :||    |:  | |       | |     |  :|||||:|  | ||| || | ||||||:
Db        1 MEG---DGSDLEPPDAGDDSKSENGENAP--IYCICRKPDINCFMIGCDNCNEWFHGDCI 55

Qy      295 GISEARGRLLERNGEDYICPNC----TILQVQ----------DETNGSATNEQDSGCRS 339
            |:|   : :     :: |  |      |:::           |  ||   ::  | :
Db       56 RITEKMAKAI----REWYCRECREKDPKLEIRYRHKKCRERDGSERAGSEPRDEGGGRKR 111

Qy      340 VGAD---------GT---------------------------------------------- 345
              :|         ||
Db      112 PASDPELQRRAGSGTGVGAMLARGSASPHKSSPQPLVATPSQHHHQQQQQQQQQIKRSAR 171

Qy      346 ------------DC-----------------------------------------TSIGTV- 353
                        ||                                         :|:  |
Db      172 MCGECEACRRTEDCGHCDFCRDMKKFGGPNKIRQKCRLRQCQLRARESYKYFPSSLSPVT 231

Qy      354 --------------------EQKSG---EDQG------IKGRIEKAANP----------- 373
                               || |   ||:|      :|   | | |
Db      232 PSEALPRPRRPPPTQQQPQQSQKLGRIREDEGTVLSSVVKEPPEATATPEPLSDEDLALD 291

Qy      374 ------------------------------------------------SGKKKLKIFQ- 383
                                                             | ||| : ::
Db      292 PDLYQDFCAGAFDDHGLPWMSDAEESPFLDPALRKRAVKVKHVKRREKKSEKKKEERYKR 351

Qy      384 -----------------PVVEAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAA 420
                              :    |:|:||||   ||| | |||:|| :| ||
Db      352 HRQKQKHKDKWKHPERADAKDPASLPQCLGPGCVRAAQPGSKYCSDDCGMKLAA 405
```

```
RESULT 12
CGBP_HUMAN
ID   CGBP_HUMAN     STANDARD;      PRT;   656 AA.
AC   Q9P0U4; Q96BC8; Q9P2V7;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   CpG binding protein (Protein containing PHD finger and CXXC domain 1).
GN   CGBP OR PCCX1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20153771; PubMed=10688657;
RA   Shin Voo K., Carlone D.L., Jacobsen B.M., Flodin A., Skalnik D.G.;
RT   "Cloning of a mammalian transcriptional activator that binds
RT   unmethylated CpG motifs and shares a CXXC domain with DNA
RT   methyltransferase, human trithorax, and methyl-CpG binding domain
RT   protein 1.";
RL   Mol. Cell. Biol. 20:2108-2121(2000).
RN   [2]
RP   SEQUENCE FROM N.A., AND DNA-BINDING.
RX   MEDLINE=20261256; PubMed=10799292;
RA   Fujino T., Hasegawa M., Shibata S., Kishimoto T., Imai S.-I.,
RA   Takano T.;
RT   "PCCX1, a novel DNA-binding protein with PHD finger and CXXC domain,
RT   is regulated by proteolysis.";
RL   Biochem. Biophys. Res. Commun. 271:305-310(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=21154917; PubMed=11230166;
RA   Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA   Ansorge W., Boecher M., Bloecker H., Bauersachs S., Blum H.,
RA   Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
RA   Mewes H.-W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
RA   Wambutt R., Korn B., Klein M., Poustka A.;
RT   "Towards a catalog of human genes and proteins: sequencing and
RT   analysis of 500 novel complete protein coding human cDNAs.";
RL   Genome Res. 11:422-435(2001).
RN   [4]
RP   SEQUENCE FROM N.A.
```

```
RC   TISSUE=Colon, and Skin;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [5]
RP   DNA-BINDING DOMAIN.
RX   MEDLINE=21576202; PubMed=11572867;
RA   Lee J.H., Voo K.S., Skalnik D.G.;
RT   "Identification and characterization of the DNA binding domain of
RT   CpG-binding protein.";
RL   J. Biol. Chem. 276:44669-44676(2001).
CC   -!- FUNCTION: Transcriptional activator that exhibits a unique DNA
CC       binding specificity for [AC]CpG[AC] unmethylated CpG motifs.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: Ubiquitous.
CC   -!- DOMAIN: The acidic domain carries the potential to activate
CC       transcription.
CC   -!- PTM: May be regulated by proteolysis.
CC   -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC   -!- SIMILARITY: Contains 1 CXXC-type zinc finger.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF149758; AAF37799.1; -.
DR   EMBL; AB031069; BAA96307.1; -.
DR   EMBL; AL136862; CAB66796.1; -.
DR   EMBL; BC014940; AAH14940.1; -.
DR   EMBL; BC015733; AAH15733.1; -.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0016563; F:transcriptional activator activity; IDA.
DR   GO; GO:0045322; F:unmethylated CpG binding; IDA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IDA.
DR   InterPro; IPR002857; Znf_CXXC.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF00628; PHD; 1.
DR   Pfam; PF02008; zf-CXXC; 1.
DR   SMART; SM00249; PHD; 1.
DR   PROSITE; PS01359; ZF_PHD_1; 1.
DR   PROSITE; PS50016; ZF_PHD_2; 1.
KW   Transcription regulation; Activator; DNA-binding; Zinc-finger; Zinc;
KW   Metal-binding; Coiled coil; Nuclear protein.
FT   ZN_FING      28     76       PHD-TYPE.
FT   ZN_FING     161    208       CXXC-TYPE.
FT   DOMAIN      256    317       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      321    359       ARG/LYS-RICH (BASIC).
FT   DOMAIN      422    474       COILED COIL (POTENTIAL).
FT   DOMAIN      156    161       POLY-GLN.
FT   DNA_BIND    162    221
FT   CONFLICT    117    117       D -> N (IN REF. 1).
FT   CONFLICT    302    302       H -> N (IN REF. 1).
FT   CONFLICT    340    340       K -> KVMER (IN REF. 4; AAH15733).
SQ   SEQUENCE   656 AA;  75711 MW;  6D2376E449905A18 CRC64;

  Query Match             5.2%;  Score 166.5;  DB 1;  Length 656;
  Best Local Similarity   18.4%;  Pred. No. 0.043;
  Matches   74;  Conservative   31;  Mismatches   55;  Indels  243;  Gaps   13;

Qy        245 EENPREAGKPKPECEVYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEARGRLL 304
              |::  | |:  |          :|||||:|  | ||| || | |||||||: |:|   : :
Db         15 EDSKSENGENAP---------IYCICRKPDINCFMIGCDNCNEWFHGDCIRITEKMAKAI 65

Qy        305 ERNGEDYICPNC----TILQVQ------DETNGSATNEQDS------------------- 335
                 :: |  |       |:::       | :|   ||:||
Db         66 ----REWYCRECREKDPKLEIRYRHKKSRERDG---NERDSSEPRDEGGGRKRPVPDPDL 118

Qy        336 --------GCRSVGADGT---------------------------------------- 345
                     |  :: | |:
Db        119 QRRAGSGTGVGAMLARGSASPHKSSPQPLVATPSQHHQQQQQQIKRSARMCGECEACRRT 178

Qy        346 -DC-----------------------------------------TSIGTV---------- 353
               ||                                       :|:  |
Db        179 EDCGHCDFCRDMKKPGGPNKIRQKCRLRQCQLRARESYKYFPSSLSPVTPSESLPRPRRP 238

Qy        354 ---------EQKSG---EDQG------IKGRIEKAANP---------------------- 373
                      || |   ||:|      :|   |  | |
Db        239 LPTQQQPQPSQKLGRIREDEGAVASSTVKEPPEATATPEPLSDEDLPLDPDLYQDFCAGA 298

Qy        374 -------------------------------------SGKKKLKIFQ------------ 383
                                                    | ||| : ::
Db        299 FDDHGLPWMSDTEESPFLDPALRKRAVKVKHVKRREKKSEKKKEERYKRHRQKQKHKDKW 358

Qy        384 ------PVVEAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAA 420
                       :   |:|:||||   ||| | |||:|| :| ||
Db        359 KHPERADAKDPASLPQCLGPGCVRPAQPSSKYCSDDCGMKLAA 401


RESULT 13
DMP1_RAT
ID   DMP1_RAT       STANDARD;      PRT;   489 AA.
AC   P98193;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Dentin matrix acidic phosphoprotein 1 precursor (Dentin matrix
DE   protein-1) (DMP-1) (AG1).
GN   DMP1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Tooth;
RX   MEDLINE=93286101; PubMed=8509401;
RA   George A., Sabsay B., Simonian P.A., Veis A.;
RT   "Characterization of a novel dentin matrix acidic phosphoprotein.
RT   Implications for induction of biomineralization.";
RL   J. Biol. Chem. 268:12624-12630(1993).
CC   -!- FUNCTION: Could be involved in the induction of mineralization of
CC       extracellular matrix.
CC   -!- TISSUE SPECIFICITY: Expressed in tooth particularly in odontoblast
CC       and ameloblast.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L11354; -; NOT_ANNOTATED_CDS.
```

```
KW   Biomineralization; Extracellular matrix; Signal.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17    489       DENTIN MATRIX ACIDIC PHOSPHOPROTEIN 1.
FT   DOMAIN       41     44       POLY-PRO.
FT   CARBOHYD    340    340       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    378    378       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    443    443       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   SITE        334    336       CELL ATTACHMENT SITE (POTENTIAL).
SQ   SEQUENCE   489 AA;  53058 MW;  59F8381479DDA085 CRC64;

  Query Match             5.1%;  Score 165.5;  DB 1;  Length 489;
  Best Local Similarity   26.0%;  Pred. No. 0.035;
  Matches   72;  Conservative   45;  Mismatches   97;  Indels   63;  Gaps   12;

Qy          3 DKGH-------LSNEEAPKAIKPTSK---------EFRKTWGFRRTTIAKREGAGDTEAD 46
              |:||       : :||:    :|||           ||     |||: ::: :  |:  ||
Db        205 DRGHTRMSSAGIRSEESKGDHEPTSTQDSDDSQDVEFSSRKSFRRSRVSEEDDRGEL-AD 263

Qy         47 PSEQQPQQHNLSLRRSGRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTVTDVE 106
               : ::  |  :    || ::  |:|  |:   |   : ::: |   : ||| :       |
Db        264 SNSRETQSVSTEDFRS-KEESRSETQED---TAETQSQEDSPEGQDPSSESS-------E 312

Qy        107 TASEGSVESSSSEIRSGPVSDSLGKEHPASSE-----KAKGGEEE-------------EDT 148
              | | | ||||| : |  |:| |     :|:      ::  ||:                |:
Db        313 EAGEPSQESSSSESQEGVASESRGDNPDNTSQTGDQRDSESSEEDRLNTFSSSESQSTEEQ 372

Qy        149 SDSDS-DGLTLKELQNRLRRKREQEPVERSLRGSQNRLRKK--RREEDSAETGSVQIGSA 205
              ||:| : |:| |            ||  :     ||  |: :   ||  | |: | |   :
Db        373 GDSESNESLSLSE--------ESQESAQDEDSSSQEGLQSQSASRESRSQESQSEQDSRS 424

Qy        206 EQDRPLCKQ------EPEASQGPVSQSETDDIENQLE 236
              |::|    |      |   | |  | || |:    :|
Db        425 EENRDSDSQDSSRSKEESNSTGSTSSSEEDNHPKNIE 461


RESULT 14
AK12_HUMAN
ID   AK12_HUMAN     STANDARD;      PRT;  1781 AA.
AC   Q02952; O00310; O00498; Q99970;
DT   01-FEB-1994 (Rel. 28, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   A-kinase anchor protein 12 (A-kinase anchor protein 250 kDa) (AKAP
DE   250) (Myasthenia gravis autoantigen gravin).
GN   AKAP12 OR AKAP250.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Heart;
RX   MEDLINE=97153077; PubMed=9000000;
RA   Nauert J.B., Klauck T.M., Langeberg L.K., Scott J.D.;
RT   "Gravin, an autoantigen recognized by serum from myasthenia gravis
RT   patients, is a kinase scaffold protein.";
RL   Curr. Biol. 7:52-62(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Umbilical vein endothelial cells;
RX   MEDLINE=98269042; PubMed=9604001;
RA   Sato N., Kokame K., Shimokado K., Kato H., Miyata T.;
RT   "Changes of gene expression by lysophosphatidylcholine in vascular
RT   endothelial cells: 12 up-regulated distinct genes including 5 cell
RT   growth-related, 3 thrombosis-related, and 4 others.";
RL   J. Biochem. 123:1119-1126(1998).
RN   [3]
RP   SEQUENCE OF 43-1781 FROM N.A.
RC   TISSUE=Umbilical vein endothelial cells;
RA   Bowditch R.D., Ginsberg M.H.;
RT   "Sequence of gravin cDNA isolated from a human umbilical vein
RT   endothelial cell library.";
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1477-1781 FROM N.A.
RC   TISSUE=Umbilical vein endothelial cells;
RX   MEDLINE=92395179; PubMed=1522245;
RA   Gordon T., Grove B., Loftus J.C., O'Toole T., McMillan R.,
RA   Lindstrom J., Ginsberg M.H.;
RT   "Molecular cloning and preliminary characterization of a novel
RT   cytoplasmic antigen recognized by myasthenia gravis sera.";
RL   J. Clin. Invest. 90:992-999(1992).
CC   -!- FUNCTION: Anchoring protein that mediates the subcellular
CC       compartmentation of protein kinase A (PKA) and protein kinase C
CC       (PKC).
CC   -!- SUBUNIT: Binds to dimeric RII-alpha regulatory subunit of PKC.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic. May be part of the cortical
CC       cytoskeleton.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q02952-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q02952-2; Sequence=VSP_004110, VSP_004111;
CC         Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: Expressed in endothelial cells, cultured
CC       fibroblasts and osteosarcoma, but not in platelets, leukocytes,
CC       monocytic cell lines or peripherical blood cells.
CC   -!- INDUCTION: Activated by lysophosphatidylcholine (lysoPC).
CC   -!- DOMAIN: Polybasic regions located between residues 265 and 556 are
CC       involved in binding PKC.
CC   -!- DISEASE: ANTIBODIES TO THE C-TERMINAL OF GRAVIN CAN BE PRODUCED BY
CC       PATIENTS WITH MYASTHENIA GRAVIS (MG).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U81607; AAC51366.1; -.
DR   EMBL; AP001504; AAB58938.1; -.
DR   EMBL; AB003476; BAA19927.1; -.
DR   EMBL; M96322; AAA35931.1; -.
DR   Genew; HGNC:370; AKAP12.
DR   MIM; 604698; -.
DR   GO; GO:0005737; C:cytoplasm; TAS.
DR   GO; GO:0005079; F:protein kinase A anchoring activity; TAS.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; TAS.
DR   InterPro; IPR001573; PkinA_anch.
DR   Pfam; PF03832; PkinA_anch; 3.
KW   Antigen; Repeat; Alternative splicing.
FT   DOMAIN      603    633       AKAP 1.
FT   DOMAIN      752    782       AKAP 2.
FT   DOMAIN      797    827       AKAP 3.
FT   DOMAIN       98    101       POLY-GLU.
FT   DOMAIN      265    556       INVOLVED IN PKC-BINDING (PROBABLE).
FT   DOMAIN     1540   1553       RII-BINDING (PROBABLE).
FT   VARSPLIC      1     98       Missing (in isoform 2).
FT                                /FTId=VSP_004110.
FT   VARSPLIC     99    106       EEEVIVTE -> MLGTITIT (in isoform 2).
FT                                /FTId=VSP_004111.
FT   CONFLICT    117    117       E -> K (IN REF. 2).
FT   CONFLICT    142    144       NRN -> TPEI (IN REF. 2 AND 3).
FT   CONFLICT    215    215       Q -> K (IN REF. 2).
FT   CONFLICT    448    448       G -> E (IN REF. 2 AND 3).
FT   CONFLICT    694    694       R -> G (IN REF. 2 AND 3).
FT   CONFLICT    867    867       G -> S (IN REF. 2 AND 3).
FT   CONFLICT    986    986       S -> A (IN REF. 2 AND 3).
FT   CONFLICT   1530   1530       E -> EE (IN REF. 3 AND 4).
FT   CONFLICT   1581   1581       V -> M (IN REF. 4).
FT   CONFLICT   1601   1601       Q -> L (IN REF. 2).
SQ   SEQUENCE   1781 AA;  191439 MW;  BA813937379FAC0F CRC64;
```

```
  Query Match             5.1%;  Score 165.5;  DB 1;  Length 1781;
  Best Local Similarity   20.4%;  Pred. No. 0.15;
  Matches  133;  Conservative   81;  Mismatches  201;  Indels  237;  Gaps   32;

Qy          3 DKGHLSN---EEAPKA-------IKPTSKE-----FRKTW---GFRRT------------ 32
              | |   |   |: | :       :|| :     |:| :   ||: |
Db        136 DDGQEENRNIEQIPSSESNLEELTQPTESQANDIGFKKVFKFVGFKFTVKKDKTEKPDTV 195

Qy         33 ---TIAKREGAGDTEA----DPS---------EQQPQQHNLSLRRSGRQPKRTERVEEFL 76
                 |: | || |   |    |||         | :|:|               ||: || |
Db        196 QLLTVKKDEGEGAAGAGDHQDPSLGAGEAASKESEPKQ-------------STEKPEETL 242

Qy         77 TTVRRRGKKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDSLGK------ 130
                :   : : |    :| :       : |   |      |:|  : ||:   |
Db        243 KREQSHAEISPPA---ESGQAVEECKEEGEEKQEKEPSKSAESPTSPVTSETGSTFKKFF 299

Qy        131 -------EHPASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLRRKREQEPVERSLRGSQN 183
                    |  | |  |:| : |:           :|:||||
Db        300 TQGWAGWRKKTSFRKPK--EDEVEASE---------------KKKEQEP---------- 331

Qy        184 RLRKKRREED-SAETGSVQIGSAEQDRPLCKQEP-EASQGPVSQSETDDI----ENQLEG 237
                |   |||  ||  | :: ::||  |   ||| |::  |   :| : :    | |: |
Db        332 --EKVDTEEDGKAEVASEKLTASEQAHP---QEPAESAHEPRLSAEYEKVELPSEEQVSG 386

Qy        238 KATQGNTEENPREAGKPKP-ECEVYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGI 296
              :|| :||       || |   ||:|          : |   :                | :
Db        387 --SQGPSEE------KPAPLATEVFDEKI------EVHQEEVVA------------EVHV 420

Qy        297 SEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQDSGCRSVGADGTDCTSIGTVEQK 356
              |    | |:  |              :|| ||    |:  |  :   :      :  :::
Db        421 STVEERTEEQKTE------------VEETAGSVPAEELVGMDAEPQEAEPAKELVKLKET 468

Qy        357 --SGED--QGI-------------KGRIEKAANPSGKKKLKI----FQPVVEAPGAPKCI 395
                ||||  ||              :| : :    | ::::|:     : :  : |  |
Db        469 CVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERMKVQGSPLKKLFTSTGLKKLS 528

Qy        396 G---------------------------------PGCSSVAQPDSVYCSNDCILKHAAA 421
              |                                   ||  :|: : |     |:   |
Db        529 GKKQKGKRGGGDEESGEHTQVPADSPDSQEEQKGESSASSPEEPEEITC-----LEKGLA 583

Qy        422 TMRPLSSGKEQKTKPKEKVK------TKPEKFSLPKCSVQVGIKISSVHKRLASEKRENP 475
              ::     :|  |   || :         :|   ||  |          :| :   :|:
Db        584 EVQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRV----------RRPSESDKEDE 633

Qy        476 VKKVMLASRSETSGKEAACESSTPSWASDHNYNAVKPEKPEKPTALSPTLLS 527
              : ||  |: | |          || |   :    :|:  |||:|     | :|
Db        634 LDKVKSATLSST--------ESTASEMQEEMKGSVEEPKPEEPKRKVDTSVS 677


RESULT 15
UN89_CAEEL
ID   UN89_CAEEL     STANDARD;      PRT;  6632 AA.
AC   O01761; Q17362;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
GN   UNC-89 OR C09D1.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC   STRAIN=Bristol N2;
RX   MEDLINE=96180278; PubMed=8603916;
RA   Benian G.M., Tinley T.L., Tang X., Borodovsky M.;
RT   "The Caenorhabditis elegans gene unc-89, required for muscle M-line
RT   assembly, encodes a giant modular protein composed of Ig and signal
RT   transduction domains.";
RL   J. Cell Biol. 132:835-848(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Du Z., Le T.T., Wilson R.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   REVISIONS.
RA   Waterston R.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Structural component of the muscle M-line. Myofilament
CC       lattice assembly begins with positional cues laid down in the
CC       basement membrane and muscle cell membrane. UNC-89 responds to
CC       these signals, localizes, and then participates in assembling an
CC       M-line.
CC   -!- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
CC   -!- SIMILARITY: Contains 1 DBL-homology (DH) domain.
CC   -!- SIMILARITY: Contains 1 fibronectin type III domain.
CC   -!- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 1 PH domain.
CC   -!- SIMILARITY: Contains 5 RCSD domains.
CC   -!- SIMILARITY: Contains 1 SH3 domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U33058; AAB00542.1; -.
DR   EMBL; AF003131; AAB54132.2; -.
DR   PDB; 1FHO; 20-DEC-00.
DR   WormPep; C09D1.1; CE30426.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR007850; RCSD.
DR   InterPro; IPR000219; RhoGEF.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00041; fn3; 1.
DR   Pfam; PF00047; ig; 47.
DR   Pfam; PF00169; PH; 1.
DR   Pfam; PF05177; RCSD; 5.
DR   Pfam; PF00621; RhoGEF; 1.
DR   Pfam; PF00018; SH3; 1.
DR   SMART; SM00408; IGc2; 23.
DR   SMART; SM00325; RhoGEF; 1.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50010; DH_2; 1.
DR   PROSITE; PS50835; IG_LIKE; 49.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;
KW   3D-structure.
FT   DOMAIN       63    127       SH3.
FT   DOMAIN      152    330       DH.
FT   DOMAIN      342    498       PH.
FT   DOMAIN      547    633       IG-LIKE C2-TYPE 1.
FT   DOMAIN      648    736       IG-LIKE C2-TYPE 2.
FT   DOMAIN      748    838       IG-LIKE C2-TYPE 3.
FT   DOMAIN      946   1033       IG-LIKE C2-TYPE 4.
FT   DOMAIN     1044   1132       IG-LIKE C2-TYPE 5.
FT   DOMAIN     1140   1227       IG-LIKE C2-TYPE 6.
FT   DOMAIN     1272   1315       THR-RICH.
FT   DOMAIN     1375   1475       RCSD 1.
FT   DOMAIN     1479   1585       RCSD 2.
FT   DOMAIN     1597   1695       RCSD 3.
FT   DOMAIN     1700   1799       RCSD 4.
FT   DOMAIN     1800   1860       RCSD 5.
FT   DOMAIN     1982   2067       IG-LIKE C2-TYPE 7.
FT   DOMAIN     2071   2163       IG-LIKE C2-TYPE 8.
```

```
FT   DOMAIN      2171   2261       IG-LIKE C2-TYPE 9.
FT   DOMAIN      2269   2359       IG-LIKE C2-TYPE 10.
FT   DOMAIN      2367   2455       IG-LIKE C2-TYPE 11.
FT   DOMAIN      2463   2564       IG-LIKE C2-TYPE 12.
FT   DOMAIN      2563   2651       IG-LIKE C2-TYPE 13.
FT   DOMAIN      2657   2746       IG-LIKE C2-TYPE 14.
FT   DOMAIN      2754   2858       IG-LIKE C2-TYPE 15.
FT   DOMAIN      2887   2980       IG-LIKE C2-TYPE 16.
FT   DOMAIN      2994   3081       IG-LIKE C2-TYPE 17.
FT   DOMAIN      3087   3183       IG-LIKE C2-TYPE 18.
FT   DOMAIN      3189   3280       IG-LIKE C2-TYPE 19.
FT   DOMAIN      3286   3376       IG-LIKE C2-TYPE 20.
FT   DOMAIN      3384   3469       IG-LIKE C2-TYPE 21.
FT   DOMAIN      3482   3572       IG-LIKE C2-TYPE 22.
FT   DOMAIN      3580   3667       IG-LIKE C2-TYPE 23.
FT   DOMAIN      3686   3777       IG-LIKE C2-TYPE 24.
FT   DOMAIN      3817   3908       IG-LIKE C2-TYPE 25.
FT   DOMAIN      3920   4009       IG-LIKE C2-TYPE 26.
FT   DOMAIN      4018   4106       IG-LIKE C2-TYPE 27.
FT   DOMAIN      4109   4201       IG-LIKE C2-TYPE 28.
FT   DOMAIN      4212   4297       IG-LIKE C2-TYPE 29.
FT   DOMAIN      4302   4387       IG-LIKE C2-TYPE 30.
FT   DOMAIN      4400   4485       IG-LIKE C2-TYPE 31.
FT   DOMAIN      4489   4580       IG-LIKE C2-TYPE 32.
FT   DOMAIN      4588   4678       IG-LIKE C2-TYPE 33.
FT   DOMAIN      4681   4771       IG-LIKE C2-TYPE 34.
FT   DOMAIN      4873   4961       IG-LIKE C2-TYPE 35.
FT   DOMAIN      4965   5057       IG-LIKE C2-TYPE 36.
FT   DOMAIN      5067   5160       IG-LIKE C2-TYPE 37.
FT   DOMAIN      5171   5260       IG-LIKE C2-TYPE 38.
FT   DOMAIN      5277   5366       IG-LIKE C2-TYPE 39.
FT   DOMAIN      5383   5472       IG-LIKE C2-TYPE 40.
FT   DOMAIN      5487   5578       IG-LIKE C2-TYPE 41.
FT   DOMAIN      5595   5685       IG-LIKE C2-TYPE 42.
FT   DOMAIN      5701   5790       IG-LIKE C2-TYPE 43.
FT   DOMAIN      5815   5904       IG-LIKE C2-TYPE 44.
FT   DOMAIN      5925   6014       IG-LIKE C2-TYPE 45.
FT   DOMAIN      6038   6130       IG-LIKE C2-TYPE 46.
FT   DOMAIN      6150   6239       IG-LIKE C2-TYPE 47.
FT   DOMAIN      6275   6368       FIBRONECTIN TYPE-III.
FT   DOMAIN      6413   6502       IG-LIKE C2-TYPE 48.
FT   DOMAIN      6507   6596       IG-LIKE C2-TYPE 49.
FT   DISULFID     568    621       POTENTIAL.
FT   DISULFID    2908   2975       POTENTIAL.
FT   DISULFID    3015   3065       POTENTIAL.
FT   DISULFID    3707   3759       POTENTIAL.
FT   DISULFID    3826   3890       POTENTIAL.
FT   DISULFID    5092   5157       POTENTIAL.
FT   DISULFID    5298   5350       POTENTIAL.
FT   DISULFID    5508   5560       POTENTIAL.
FT   DISULFID    5616   5669       POTENTIAL.
FT   DISULFID    5722   5764       POTENTIAL.
FT   DISULFID    5836   5901       POTENTIAL.
FT   DISULFID    5946   5998       POTENTIAL.
FT   DISULFID    6036   6171       POTENTIAL.
FT   DISULFID    6421   6486       POTENTIAL.
FT   CONFLICT    2137   2137       A -> P (IN REF. 1).
FT   CONFLICT    2245   2247       AKA -> PKP (IN REF. 1).
FT   CONFLICT    2258   2258       A -> P (IN REF. 1).
FT   CONFLICT    2284   2284       E -> G (IN REF. 1).
FT   CONFLICT    2297   2297       M -> I (IN REF. 1).
FT   CONFLICT    3531   3531       A -> G (IN REF. 1).
FT   CONFLICT    3884   3888       DAGEY -> RRRRI (IN REF. 1).
FT   CONFLICT    3929   3929       A -> V (IN REF. 1).
FT   CONFLICT    5134   5134       A -> P (IN REF. 1).
FT   CONFLICT    5145   5145       T -> S (IN REF. 1).
FT   CONFLICT    5185   5185       G -> A (IN REF. 1).
FT   CONFLICT    5199   5199       K -> N (IN REF. 1).
FT   CONFLICT    5202   5202       L -> F (IN REF. 1).
FT   CONFLICT    5213   5213       F -> L (IN REF. 1).
FT   CONFLICT    6178   6178       A -> G (IN REF. 1).
FT   CONFLICT    6268   6268       K -> E (IN REF. 1).
SQ   SEQUENCE   6632 AA;  731665 MW;  262D3EDD62960E89 CRC64;
```

```
  Query Match             5.1%;  Score 165;  DB 1;  Length 6632;
  Best Local Similarity   19.9%;  Pred. No. 0.65;
  Matches  123;  Conservative   87;  Mismatches  214;  Indels  194;  Gaps   24;

Qy          8 SNEEAPKAIKPTSKEPRKTWGFRRTTIAKREGAGDTE---ADPSEQQPQQHNLSLRRSGR 64
              : || || :  : :: :|:     :   | :    ||   | |:::  ::      :|
Db       1358 AEEELPKEVIDSDRKKKKS----PSPDKKEKSPEKTEEKPASPTKKTGEEVKSPKEKSPA 1413

Qy         65 QPKRTER---VEEFLTTVRRRGKKNVPVSLEDS-SEPTSSTVTDVE--------TASEGS 112
               | : |:    ||  :  ::     : |   | | | ||  |  :|:         | | |
Db       1414 SPTKKEKSPAAEEVKSPTKKEKSPSSPTKKEKSPSSPTKKTGDEVKEKSPPKSPTKKEKS 1473

Qy        113 ----------------------VESSSEIRSGPVSDSLGKEHPASSEKAKGGEEEEDT-- 148
                                   || |||        ::  |    ||:::   ::| |
Db       1474 PEKPEDVKSPVKKEKSPDATNIVEVSSETTIEKTETTMTTEMTHESEESRTSVKKEKTPE 1533

Qy        149 ----------SDSDSDGLTLKELQNRLRRKREQEPVERS----------------LRG 180
                         |   :    :|::: :::::  | ||                  :
Db       1534 KVDEKPKSPTKKDKSPEKSITEEIKSPVKKEKSPEKVEEKPASPTKKEKSPEKPASPTKK 1593

Qy        181 SQNRLRKKRREEDSAETGSVQIGSAEQDRPLCKQEPEASQGPVS---------QSETDDI 231
              |:| ::   ::| | |   |:    |  | |  :|  | |          :| |:|:
Db       1594 SENEVKSPTKKEKSPEKSVVE----ELKSPKEKSPEKADDKPKSPTKKEKSPEKSATEDV 1649

Qy        232 ENQLEGKATQGNTEENPREAGK------PKPECEVYDPNALYCICRQPHNNRFMICCDRC 285
              ::  : : :    || |    |       | : ||  |       : |
Db       1650 KSPTKKEKSPEKVEEKPTSPTKKESSPTKKTDDEVKSPTKKE---KSP------------ 1694

Qy        286 EEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQDSGCRSVGADGT 345
                                              |  :|   | | :: |  :||
Db       1695 ---------------------------------QTVEEKPASPTKKEKSPEKSV----- 1715

Qy        346 DCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKL--KIFQPVVEAPGAPKCIGPGCSSVA 403
                  | || :::  :   ||  :|: |:|   |      |::|   |   |  |:
Db       1716 ------VEEVKSPKEKSPEKAEEKPKSPTKKEKSPEKSAAEEVKSP-TKKEKSPEKSAEE 1768

Qy        404 QPDSVYCSNDCILKHAAATMRFLSSGKEQKTKPKEKVKTKPEKFSLPKCSVQVGIKISSV 463
              :| |         :| |   ::       :| |  |||: ||              |
Db       1769 KPKSPTKKESSPVKMADDEVK----SPTKKEKSPEKVEEKPAS----------PTKKEKT 1814

Qy        464 HKRLASEKRENPVKKVMLASRSETS-GKEAACESSTPSWASDHNYNAVKPEKP-EKPTAL 521
               :: |:|: ::| ||     :| :|  |:   ||   |            |||| |||
Db       1815 PEKSAAEELKSPTKK----EKSPSSPTKKTGDESKEKS----------PEKPEEKPK-- 1857

Qy        522 SPTLLSKCTYHPKAGFPG 539
              |||         ||  ||
Db       1858 SPT--------PKKSPPG 1867


Search completed: April 20, 2004, 21:18:53
Job time : 21 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 20, 2004, 21:19:55 ; Search time 48 Seconds
(without alignments)
3526.319 Million cell updates/sec

Title: US-09-787-016A-4
Perfect score: 3228
Sequence: 1 MDDKGHLSNEEAPKAIKPTS.....LKSCVGLMLEAISYFSFRPW 614

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1124875 seqs, 275673149 residues

Total number of hits satisfying chosen parameters: 1124875

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	740	22.9	181	9	US-09-925-301-1169	Sequence 1169, Ap
2	194.5	6.0	676	10	US-09-823-187-24	Sequence 24, Appl
3	194.5	6.0	676	10	US-09-863-776-14	Sequence 14, Appl
4	191.5	5.9	676	14	US-10-203-708-28	Sequence 28, Appl
5	191.5	5.9	717	14	US-10-203-708-29	Sequence 29, Appl
6	190.5	5.9	518	9	US-09-925-299-862	Sequence 862, App
7	190.5	5.9	518	10	US-09-925-299-862	Sequence 862, App
8	190.5	5.9	667	10	US-09-863-776-16	Sequence 16, Appl
9	190.5	5.9	667	10	US-09-863-776-18	Sequence 18, Appl
10	188.5	5.8	555	14	US-10-106-698-5847	Sequence 5847, Ap
11	188.5	5.8	670	10	US-09-823-187-86	Sequence 86, Appl
12	188.5	5.8	670	10	US-09-863-776-51	Sequence 51, Appl
13	188.5	5.8	671	10	US-09-946-374-308	Sequence 308, App
14	188.5	5.8	671	10	US-09-823-187-85	Sequence 85, Appl
15	188.5	5.8	671	10	US-09-863-776-20	Sequence 20, Appl
16	188.5	5.8	671	12	US-10-206-915-346	Sequence 346, App
17	188.5	5.8	671	12	US-10-199-670-346	Sequence 346, App
18	188.5	5.8	671	12	US-10-201-858-346	Sequence 346, App
19	188.5	5.8	671	12	US-10-205-890-346	Sequence 346, App
20	188.5	5.8	671	12	US-10-208-024-346	Sequence 346, App
21	188.5	5.8	671	12	US-10-201-853-346	Sequence 346, App
22	188.5	5.8	671	12	US-10-174-581-346	Sequence 346, App
23	188.5	5.8	671	12	US-10-176-483-346	Sequence 346, App
24	188.5	5.8	671	12	US-10-176-749-346	Sequence 346, App
25	188.5	5.8	671	12	US-10-176-914-346	Sequence 346, App
26	188.5	5.8	671	12	US-10-176-915-346	Sequence 346, App
27	188.5	5.8	671	12	US-10-006-485A-308	Sequence 308, App
28	188.5	5.8	671	12	US-10-013-907A-308	Sequence 308, App
29	188.5	5.8	671	12	US-10-015-499A-308	Sequence 308, App
30	188.5	5.8	671	12	US-10-176-484-346	Sequence 346, App
31	188.5	5.8	671	12	US-10-180-550-346	Sequence 346, App
32	188.5	5.8	671	12	US-10-183-014-346	Sequence 346, App
33	188.5	5.8	671	12	US-10-187-738-346	Sequence 346, App
34	188.5	5.8	671	12	US-10-187-740-346	Sequence 346, App
35	188.5	5.8	671	12	US-10-187-883-346	Sequence 346, App
36	188.5	5.8	671	12	US-10-194-363-346	Sequence 346, App
37	188.5	5.8	671	12	US-10-194-460-346	Sequence 346, App
38	188.5	5.8	671	12	US-10-194-463-346	Sequence 346, App
39	188.5	5.8	671	12	US-10-194-484-346	Sequence 346, App
40	188.5	5.8	671	12	US-10-195-884-346	Sequence 346, App
41	188.5	5.8	671	12	US-10-195-896-346	Sequence 346, App
42	188.5	5.8	671	12	US-10-196-744-346	Sequence 346, App
43	188.5	5.8	671	12	US-10-196-755-346	Sequence 346, App
44	188.5	5.8	671	12	US-10-196-757-346	Sequence 346, App
45	188.5	5.8	671	12	US-10-197-704-346	Sequence 346, App

ALIGNMENTS

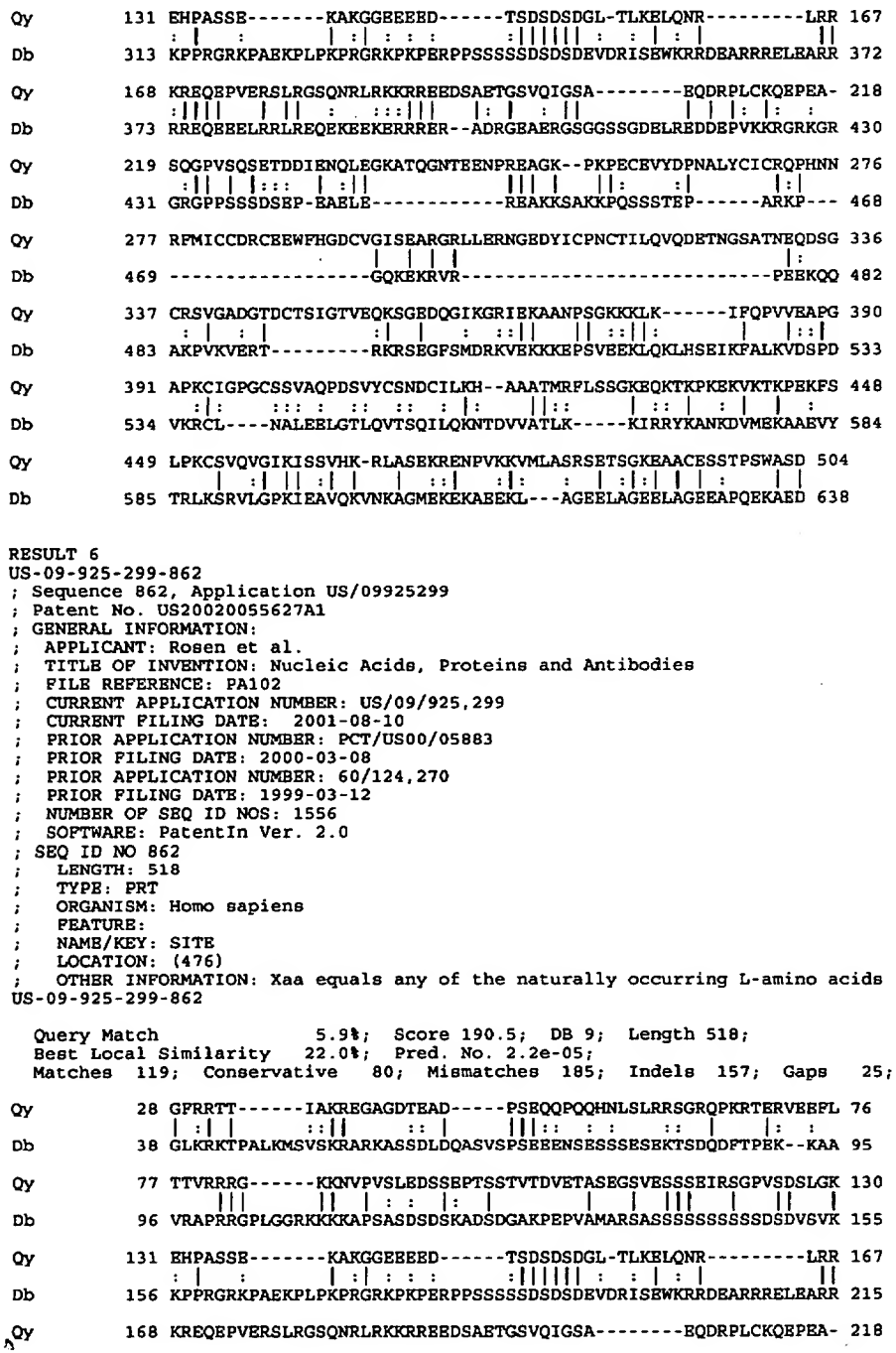
RESULT 1
US-09-925-301-1169
; Sequence 1169, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1169
; LENGTH: 181
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-301-1169

Query Match 22.9%; Score 740; DB 9; Length 181;
Best Local Similarity 85.1%; Pred. No. 3.2e-46;
Matches 131; Conservative 8; Mismatches 15; Indels 0; Gaps 0;

```
Qy      252 GKPKPECEVYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDY 311
            |:|||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       11 GRPKPECEGYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDY 70

Qy      312 ICPNCTILQVQDETNGSATNEQDSGCRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAA 371
            |||||||||||||:    ::|::  |  ||||||||||:|||| ||||||||||||||
Db       71 ICPNCTILQVQDETHSETADQQEAKWRPGDADGTDCTSIGTIEQKSSEDQGIKGRIEKAA 130

Qy      372 NPSGKKKLKIFQPVVEAPGAPKCIGPGCSSVAQP 405
            |||||||||||||:||||| |||||||  || |
Db      131 NPSGKKKLKIFQPVIEAPGASKCIGPGCCHVAHP 164
```

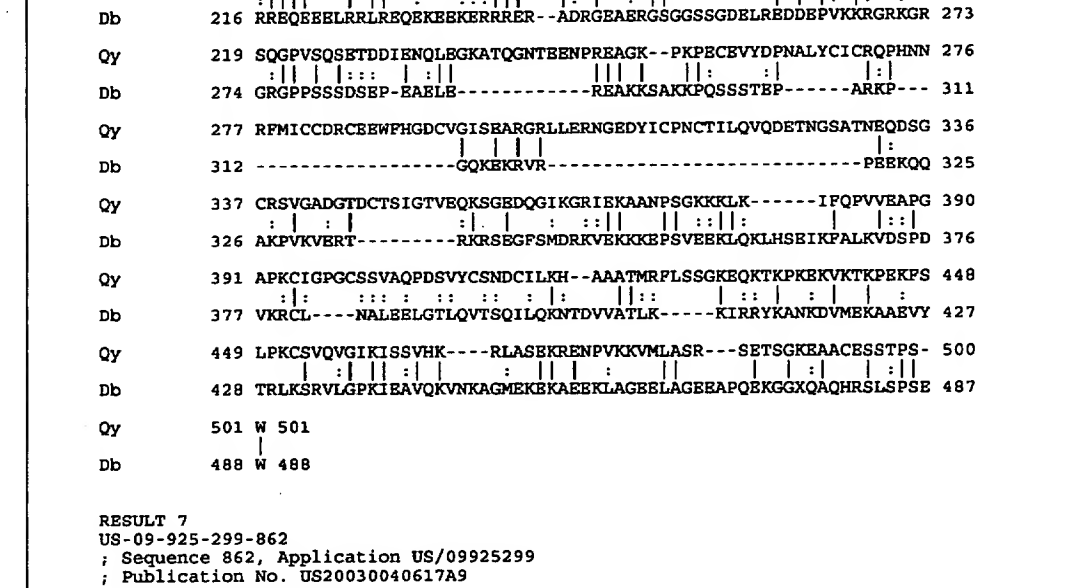
```
RESULT 2
US-09-823-187-24
; Sequence 24, Application US/09823187
; Publication No. US20030096952A1
; GENERAL INFORMATION:
;  APPLICANT: Burgess, Catherine
;  APPLICANT:  Gusev, Vladimir Y
;  APPLICANT:  Liu, Xiaohong
;  APPLICANT:  Majumder, Kumud
;  APPLICANT:  Padigaru, Muralidhar
;  APPLICANT:  Patturajan, Meera
;  APPLICANT:  Shimkets, Richard A
;  APPLICANT:  Spaderna, Steven K
;  APPLICANT:  Spytek, Kimberly
;  APPLICANT:  Taupier, Raymond J
;  TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
;  FILE REFERENCE: 15966-745
;  CURRENT APPLICATION NUMBER: US/09/823,187
;  CURRENT FILING DATE:  2001-03-29
;  PRIOR APPLICATION NUMBER: 60/193,339
;  PRIOR FILING DATE: 2000-03-30
;  PRIOR APPLICATION NUMBER: 60/193,205
;  PRIOR FILING DATE: 2000-03-30
;  PRIOR APPLICATION NUMBER: 60/195,343
;  PRIOR FILING DATE: 2000-04-05
;  PRIOR APPLICATION NUMBER: 60/195,088
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: 60/195,005
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: 60/195,792
;  PRIOR FILING DATE: 2000-04-10
;  PRIOR APPLICATION NUMBER: 60/196,556
;  PRIOR FILING DATE: 2000-04-11
;  PRIOR APPLICATION NUMBER: 60/197,081
;  PRIOR FILING DATE: 2000-04-13
;  PRIOR APPLICATION NUMBER: 60/197,525
;  PRIOR FILING DATE: 2000-04-14
;  PRIOR APPLICATION NUMBER: 60/197,087
;  PRIOR FILING DATE: 2000-04-14
;  NUMBER OF SEQ ID NOS: 103
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
;   LENGTH: 676
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-823-187-24

  Query Match             6.0%;  Score 194.5;  DB 10;  Length 676;
  Best Local Similarity   22.0%;  Pred. No. 1.6e-05;
  Matches  118;  Conservative   83;  Mismatches  183;  Indels  153;  Gaps   24;

Qy         28 GFRRTT------IAKREGAGDTEAD-----PSEQQPQQHNLSLRRSGRQPKRTERVEEFL 76
              | :| |       ::||      :: |     |||::  : :    ::  |     |:  :
Db        154 GLKRKTPALKVSVSKRARKASSDLDQASVSPSEEENSESSSESEKTSDQDFTPEK--KAA 211

Qy         77 TTVRRRG------KKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDSLGK 130
                 |||       ||  | : :  |: |          |   |  |||   |   ||   |
Db        212 VRAPRRGPLGGRKKKKAPSASDSDSKADSDGAKPEPVAMARSASSSSSSSSSDSDVSVK 271

Qy        131 EHPASSE-------KAKGGEEEED------TSDSDSDGL-TLKELQNR---------LRR 167
              : |   :       | :| : : :      :||||||  :  : | : |          ||
Db        272 KPPRGRKPAEKPLPKPRGRKPKPERPPSSSSSDSDSDEVDRISEWKRRDEARRRELEARR 331

Qy        168 KREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSA--------EQDRPLCKQEPEA- 218
              :||||   | ||  :   :::|||   |: |  : ||         | | |: |:  :
Db        332 RREQEEELRRLREQEKEEKERRRER--ADRGEAERGSGGSSGDELREDDEPVKKRGRKGR 389

Qy        219 SQGPVSQSETDDIENQLEGKATQGNTEENPREAGK--PKPECEVYDPNALYCICRQPHNN 276
                :|| | |:::  | :||              ||| |   ||:    :|       |:|
Db        390 GRGPPSSSDSEP-EAELE-----------REAKKSAKKPQSSSTEP------ARKP--- 427

Qy        277 RFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQDSG 336
                                  | | | |                              |:
Db        428 -----------------GQKEKRVR----------------------------PEEKQQ 441

Qy        337 CRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLK------IFQPVVEAPG 390
              | | : |          :| |  :  ::||   || ::||:       |   |::|
Db        442 ARPVKVERT---------RKRSEGFSMDRKVEKKKEPSVEEKLQKLHSEIKPALKVDSPD 492

Qy        391 APKCIGPGCSSVAQPDSVYCSNDCILKH--AAATMRFLSSGKEQKTKPKEKVKTKPEKFS 448
                :|:   ::: :  ::  ::  : |:     ||::     | :: |  : |  |  :
Db        493 VKRCL----NALEELGTLQVTSQILQKNTDVVATLK-----KIRRYKANKDVMEKAAEVY 543

Qy        449 LPKCSVQVGIKISSVHK-RLASEKRENPVKKVMLASRSETSGKEAACESSTPSWASD 504
                | :| || :| |   |  ::|   :|:   :  | :|:| | | :    | |
Db        544 TRLKSRVLGPKIEAVQKVNKAGMEKEKAEEKL---AGEELAGEELAGEEAPQEKAED 597


RESULT 3
US-09-863-776-14
; Sequence 14, Application US/09863776
; Publication No. US20030198953A1
; GENERAL INFORMATION:
;  APPLICANT: Spytek, Kimberly A
;  APPLICANT:  Majumder, Kumud
;  APPLICANT:  Tchernev, Velizar T
;  APPLICANT:  Mishra, Vishnu
;  APPLICANT:  Padigaru, Muralidhara
;  APPLICANT:  Spaderna, Steven K
;  APPLICANT:  Shenoy, Suresh G
;  APPLICANT:  Rastelli, Luca
;  APPLICANT:  Li, Li
;  APPLICANT:  Taupier, Raymond J
;  APPLICANT:  Gangolli, Esha
;  TITLE OF INVENTION: No. US20030198953A1el Proteins and Nucleic Acids Encoding Same
;  FILE REFERENCE: 21402-020
;  CURRENT APPLICATION NUMBER: US/09/863,776
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: 09/540,763
;  PRIOR FILING DATE: 2000-03-30
;  PRIOR APPLICATION NUMBER: 60/206,679
;  PRIOR FILING DATE: 2000-05-24
;  PRIOR APPLICATION NUMBER: 60/206,688
;  PRIOR FILING DATE: 2000-05-24
;  PRIOR APPLICATION NUMBER: 60/206,829
;  PRIOR FILING DATE: 2000-05-24
;  PRIOR APPLICATION NUMBER: 60/207,748
;  PRIOR FILING DATE: 2000-05-30
;  PRIOR APPLICATION NUMBER: 60/207,798
;  PRIOR FILING DATE: 2000-05-30
;  PRIOR APPLICATION NUMBER: 60/208,263
;  PRIOR FILING DATE: 2000-05-31
;  PRIOR APPLICATION NUMBER: 60/208,831
;  PRIOR FILING DATE: 2000-06-02
;  PRIOR APPLICATION NUMBER: 60/209,451
;  PRIOR FILING DATE: 2000-06-05
;  PRIOR APPLICATION NUMBER: 60/210,060
;  PRIOR FILING DATE: 2000-06-07
;  PRIOR APPLICATION NUMBER: 60/219,507
;  PRIOR FILING DATE: 2000-07-20
;  PRIOR APPLICATION NUMBER: 60/221,337
;  PRIOR FILING DATE: 2000-07-26
;  PRIOR APPLICATION NUMBER: 60/221,927
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: 60/263,135
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: 60/263,688
;  PRIOR FILING DATE: 2001-01-24
;  PRIOR APPLICATION NUMBER: 60/263,694
;  PRIOR FILING DATE: 2001-01-24
;  NUMBER OF SEQ ID NOS: 155
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
;   LENGTH: 676
;   TYPE: PRT
```

```
;   ORGANISM: Homo sapiens
US-09-863-776-14

  Query Match             6.0%;  Score 194.5;  DB 10;  Length 676;
  Best Local Similarity   22.0%;  Pred. No. 1.6e-05;
  Matches  118;  Conservative   83;  Mismatches  183;  Indels  153;  Gaps   24;

Qy          28 GFRRTT------IAKREGAGDTEAD-----PSEQQPQQHNLSLRRSGRQPKRTERVEEFL 76
               | :| |       ::||      :: |      |||::  : :    ::  |    |:  :
Db         154 GLKRKTPALKVSVSKRARKASSDLDQASVSPSEEENSESSSESEKTSDQDFTPEK--KAA 211

Qy          77 TTVRRRG------KKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDSLGK 130
                  |||       ||  | : :  |:  |          |   | |||   |   ||   |
Db         212 VRAPRRGPLGGRKKKKAPSASDSDSKADSDGAKPEPVAMARSASSSSSSSSSSDSDVSVK 271

Qy         131 EHPASSE-------KAKGGEEEED------TSDSDSDGL-TLKELQNR---------LRR 167
               : |   :       | :| : : :      :||||||  :  : | : |           ||
Db         272 KPPRGRKPAEKPLPKPRGRKPKPERPPSSSSSDSDSDEVDRISEWKRRDEARRRELEARR 331

Qy         168 KREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSA--------EQDRPLCKQEPEA- 218
               :||||   | ||  :   :::|||   |: |  : ||          | | |: |:  :
Db         332 RREQEEELRRLREQEKEEKERRRER--ADRGEAERGSGGSSGDELREDDEPVKKRGRKGR 389

Qy         219 SQGPVSQSETDDIENQLEGKATQGNTEENPREAGK--PKPECEVYDPNALYCICRQPHNN 276
               :|| | |:::  | :||           ||| |   ||:    :|       |:|
Db         390 GRGPPSSSDSEP-EAELE-----------REAKKSAKKPQSSSTEP------ARKP--- 427

Qy         277 RFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQDSG 336
                           | | | |                                    |:
Db         428 -----------------GQKEKRVR----------------------------PEEKQQ 441

Qy         337 CRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLK------IFQPVVEAPG 390
               | |  : |          :| |  :  ::||  || ::||:       |   |::|
Db         442 ARPVKVERT---------RKRSEGFSMDRKVEKKKEPSVEEKLQKLHSEIKFALKVDSPD 492

Qy         391 APKCIGPGCSSVAQPDSVYCSNDCILKH--AAATMRFLSSGKEQKTKPKEKVKTKPEKFS 448
               :|:    ::: :  ::  ::  : |:    ||::     | :: |  : |  |  :
Db         493 VKRCL----NALEELGTLQVTSQILQKNTDVVATLK-----KIRRYKANKDVMEKAAEVY 543

Qy         449 LPKCSVQVGIKISSVHK-RLASEKRENPVKKVMLASRSETSGKEAACESSTPSWASD 504
                  | :| || :| |   |  ::|   :|:   :  | :|:| | | :    | |
Db         544 TRLKSRVLGPKIEAVQKVNKAGMEKEKAEEKL---AGEELAGEELAGEEAPQEKAED 597


RESULT 4
US-10-203-708-28
; Sequence 28, Application US/10203708
; Publication No. US20030149238A1
; GENERAL INFORMATION:
;  APPLICANT: SMITHKLINE BEECHAM CORPORATION
;  APPLICANT:  SMITHKLINE BEECHAM p.l.c.
;  TITLE OF INVENTION: NOVEL COMPOUNDS
;  FILE REFERENCE: GP50013
;  CURRENT APPLICATION NUMBER: US/10/203,708
;  CURRENT FILING DATE:  2002-08-13
;  PRIOR APPLICATION NUMBER: PCT/US01/04703
;  PRIOR FILING DATE: 2001-02-14
;  PRIOR APPLICATION NUMBER: 60/182,172
;  PRIOR FILING DATE: 2000-02-14
;  PRIOR APPLICATION NUMBER: 60/186,084
;  PRIOR FILING DATE: 2000-02-29
;  NUMBER OF SEQ ID NOS: 46
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 28
;   LENGTH: 676
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-203-708-28

  Query Match             5.9%;  Score 191.5;  DB 14;  Length 676;
  Best Local Similarity   21.8%;  Pred. No. 2.6e-05;
  Matches  117;  Conservative   84;  Mismatches  183;  Indels  153;  Gaps   24;

Qy          28 GFRRTT------IAKREGAGDTEAD-----PSEQQPQQHNLSLRRSGRQPKRTERVEEFL 76
               | :| |       ::||      :: |      |||::  : :    ::  |    |:  :
Db         154 GLKRKTPALKMSVSKRARKASSDLDQASVSPSEEENSESSSESEKTSDQDFTPEK--KAA 211

Qy          77 TTVRRRG------KKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDSLGK 130
                  |||       ||  | : :  |:  |          |   | |||   |   ||   |
Db         212 VRAPRRGPLGGRKKKKAPSASDSDSKADSDGAKPEPVAMARSASSSSSSSSSSDSDVSVK 271

Qy         131 EHPASSE-------KAKGGEEEED------TSDSDSDGL-TLKELQNR---------LRR 167
               : |   :       | :| : : :      :||||||  :  : | : |           ||
Db         272 KPPRGRKPAEKPLPKPRGRKPKPERPPSSSSSDSDSDEVDRISEWKRRDEARRRELEARR 331

Qy         168 KREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSA--------EQDRPLCKQEPEA- 218
               :|||| | ||  :   :::|||   |: |  : ||          | | |: |:  :
Db         332 RREQEEELRRLREQEKEEKERRRER--ADRGEAERGSGGSSGDELREDDEPVKKRGRKGR 389

Qy         219 SQGPVSQSETDDIENQLEGKATQGNTEENPREAGK--PKPECEVYDPNALYCICRQPHNN 276
               :|| | |:::  | :||           ||| |   ||:    :|       |:|
Db         390 GRGPPSSSDSEP-EAELE-----------REAKKSAKKPQSSSTEP------ARKP--- 427

Qy         277 RFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQDSG 336
                           | | | |                                    |:
Db         428 -----------------GQKEKRVR----------------------------PEEKQQ 441

Qy         337 CRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLK------IFQPVVEAPG 390
               | |  : |          :| |  :  ::||  || ::||:       |   |::|
Db         442 AKPVKVERT---------RKRSEGFSMDRKVEKKKEPSVEEKLQKLHSEIKFALKVDSPD 492

Qy         391 APKCIGPGCSSVAQPDSVYCSNDCILKH--AAATMRFLSSGKEQKTKPKEKVKTKPEKFS 448
               :|:    ::: :  ::  ::  : |:    ||::     | :: |  : |  |  :
Db         493 VKRCL----NALEELGTLQVTSQILQKNTDVVATLK-----KIRRYKANKDVMEKAAEVY 543

Qy         449 LPKCSVQVGIKISSVHK-RLASEKRENPVKKVMLASRSETSGKEAACESSTPSWASD 504
                  | :| || :| |   |  ::|   :|:   :  | :|:| | | :    | |
Db         544 TRLKSRVLGPKIEAVQKVNKAGMEKEKAEEKL---AGEELAGEELAGEEAPQEKAED 597


RESULT 5
US-10-203-708-29
; Sequence 29, Application US/10203708
; Publication No. US20030149238A1
; GENERAL INFORMATION:
;  APPLICANT: SMITHKLINE BEECHAM CORPORATION
;  APPLICANT:  SMITHKLINE BEECHAM p.l.c.
;  TITLE OF INVENTION: NOVEL COMPOUNDS
;  FILE REFERENCE: GP50013
;  CURRENT APPLICATION NUMBER: US/10/203,708
;  CURRENT FILING DATE:  2002-08-13
;  PRIOR APPLICATION NUMBER: PCT/US01/04703
;  PRIOR FILING DATE: 2001-02-14
;  PRIOR APPLICATION NUMBER: 60/182,172
;  PRIOR FILING DATE: 2000-02-14
;  PRIOR APPLICATION NUMBER: 60/186,084
;  PRIOR FILING DATE: 2000-02-29
;  NUMBER OF SEQ ID NOS: 46
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 29
;   LENGTH: 717
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-203-708-29

  Query Match             5.9%;  Score 191.5;  DB 14;  Length 717;
  Best Local Similarity   21.8%;  Pred. No. 2.8e-05;
  Matches  117;  Conservative   84;  Mismatches  183;  Indels  153;  Gaps   24;

Qy          28 GFRRTT------IAKREGAGDTEAD-----PSEQQPQQHNLSLRRSGRQPKRTERVEEFL 76
               | :| |       ::||      :: |      |||::  : :    ::  |    |:  :
Db         195 GLKRKTPALKMSVSKRARKASSDLDQASVSPSEEENSESSSESEKTSDQDFTPEK--KAA 252

Qy          77 TTVRRRG------KKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDSLGK 130
                  |||       ||  | : :  |:  |          |   | |||   |   ||   |
Db         253 VRAPRRGPLGGRKKKKAPSASDSDSKADSDGAKPEPVAMARSASSSSSSSSSSDSDVSVK 312
```



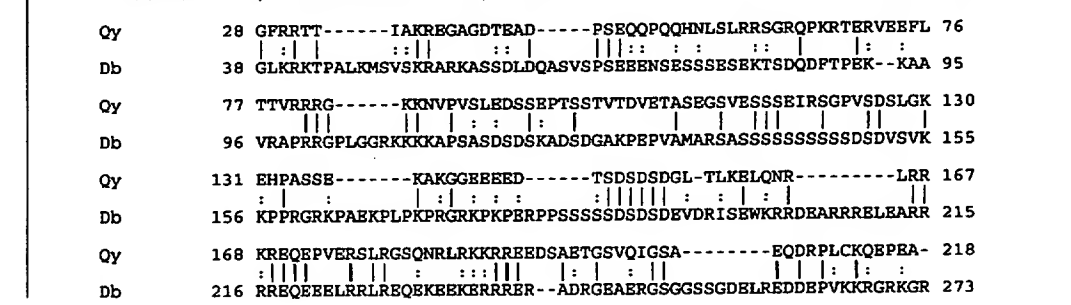
RESULT 6
US-09-925-299-862
; Sequence 862, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 862
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (476)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-862

Query Match 5.9%; Score 190.5; DB 9; Length 518;
Best Local Similarity 22.0%; Pred. No. 2.2e-05;
Matches 119; Conservative 80; Mismatches 185; Indels 157; Gaps 25;



RESULT 7
US-09-925-299-862
; Sequence 862, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 862
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (476)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-862

Query Match 5.9%; Score 190.5; DB 10; Length 518;
Best Local Similarity 22.0%; Pred. No. 2.2e-05;
Matches 119; Conservative 80; Mismatches 185; Indels 157; Gaps 25;



```
Qy        219 SQGPVSQSETDDIENQLEGKATQGNTEENPREAGK--PKPECEVYDPNALYCICRQPHNN 276
Db        274 GRGPPSSSDSEP-EAELE-----------REAKKSAKKPQSSSTEP------ARKP--- 311

Qy        277 RPMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQDSG 336
Db        312 -----------------GQKEKRVR----------------------------PEEKQQ 325

Qy        337 CRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLK------IFQPVVEAPG 390
Db        326 AKPVKVERT---------RKRSEGFSMDRKVEKKKEPSVEEKLQKLHSEIKFALKVDSPD 376

Qy        391 APKCIGPGCSSVAQPDSVYCSNDCILKH--AAATMRFLSSGKEQKTKPKEKVKTKPEKFS 448
Db        377 VKRCL----NALEELGTLQVTSQILQKNTDVVATLK-----KIRRYKANKDVMEKAAEVY 427

Qy        449 LPKCSVQVGIKISSVHK----RLASEKRENPVKKVMLASR---SETSGKEAACESSTPS- 500
Db        428 TRLKSRVLGPKIEAVQKVNKAGMEKEKAEEKLAGEELAGEEAPQEKGGXQAQHRSLSPSE 487

Qy        501 W 501
Db        488 W 488
```

RESULT 8
US-09-863-776-16
; Sequence 16, Application US/09863776
; Publication No. US20030198953A1
; GENERAL INFORMATION:
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Majumder, Kumud
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Mishra, Vishnu
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Spaderna, Steven K
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Rastelli, Luca
; APPLICANT: Li, Li
; APPLICANT: Taupier, Raymond J
; APPLICANT: Gangolli, Esha
; TITLE OF INVENTION: No. US20030198953A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-020
; CURRENT APPLICATION NUMBER: US/09/863,776
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 09/540,763
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/206,679
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/206,688
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/206,829
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/207,748
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/207,798
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/208,263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 60/208,831
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/209,451
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: 60/210,060
; PRIOR FILING DATE: 2000-06-07
; PRIOR APPLICATION NUMBER: 60/219,507
; PRIOR FILING DATE: 2000-07-20
; PRIOR APPLICATION NUMBER: 60/221,337
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/221,927
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 60/263,135
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/263,688
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: 60/263,694
; PRIOR FILING DATE: 2001-01-24
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 667
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-863-776-16

Query Match 5.9%; Score 190.5; DB 10; Length 667;
Best Local Similarity 22.5%; Pred. No. 3.1e-05;
Matches 120; Conservative 84; Mismatches 179; Indels 151; Gaps 25;

```
Qy         28 GFRRTT------IAKREGAGDTEAD-----PSEQQPQQHNLSLRRSGRQPKRTERVEEFL 76
Db        154 GLKRKTPALKVSVSKRARKASSDLDQASVSPSEEENSESSSESEKTSDQDFTPEK--KAA 211

Qy         77 TTVRRRG------KKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSSEIRSGPVSDSLGK 130
Db        212 VRAPRRGPLGGRKKKKAPSASDSDSKADSDGAKPEPVAMARSASSSSSSSSSSDSDVSVK 271

Qy        131 EHPASSE-------KAKGGEEEED------TSDSDSDGL-TLKELQNR---------LRR 167
Db        272 KPPRGRKPAEKPLPKPRGRKPKPERPPSSSSSDSDSDEVDRISEWKRRDEARRRELEARR 331

Qy        168 KREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSA-----EQDRPLCKQEPEA-SQG 221
Db        332 RREQEEELRRLR-EQEKEEKERRRERAAERGSG--GSSGDELREDDEPVKKRGRKGRGRG 388

Qy        222 PVSQSETDDIENQLEGKATQGNTEENPREAGK--PKPECEVYDPNALYCICRQPHNNRFM 279
Db        389 PPSSSDSEP-EAELE-----------REAKKSAKKPQSSSTEP------ARKP------ 423

Qy        280 ICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQDSGCRS 339
Db        424 --------------GQKEKRVR----------------------------PEEKQQAKP 440

Qy        340 VGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLK------IFQPVVEAPGAPK 393
Db        441 VKVERT---------RKRSEGFSMDRKVEKKKEPSVEEKLQKLHSEIKFALKVDSPDVKR 491

Qy        394 CIGPGCSSVAQPDSVYCSNDCILKH--AAATMRFLSSGKEQKTKPKEKVKTKPEKPSLPK 451
Db        492 CL----NALEELGTLQVTSQILQKNTDVVATLK-----KIRRYKANKDVMEKAAEVYTRL 542

Qy        452 CSVQVGIKISSVHK-RLASEKRENPVKKVMLASRSETSGKEAACESSTPSWASD 504
Db        543 KSRVLGPKIEAVQKVDKAGMEKEKAEEKL---AGEELAGEEAPQEKAEDKPSTD 593
```

RESULT 9
US-09-863-776-18
; Sequence 18, Application US/09863776
; Publication No. US20030198953A1
; GENERAL INFORMATION:
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Majumder, Kumud
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Mishra, Vishnu
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Spaderna, Steven K
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Rastelli, Luca
; APPLICANT: Li, Li
; APPLICANT: Taupier, Raymond J
; APPLICANT: Gangolli, Esha
; TITLE OF INVENTION: No. US20030198953A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-020
; CURRENT APPLICATION NUMBER: US/09/863,776

```
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: 09/540,763
;  PRIOR FILING DATE: 2000-03-30
;  PRIOR APPLICATION NUMBER: 60/206,679
;  PRIOR FILING DATE: 2000-05-24
;  PRIOR APPLICATION NUMBER: 60/206,688
;  PRIOR FILING DATE: 2000-05-24
;  PRIOR APPLICATION NUMBER: 60/206,829
;  PRIOR FILING DATE: 2000-05-24
;  PRIOR APPLICATION NUMBER: 60/207,748
;  PRIOR FILING DATE: 2000-05-30
;  PRIOR APPLICATION NUMBER: 60/207,798
;  PRIOR FILING DATE: 2000-05-30
;  PRIOR APPLICATION NUMBER: 60/208,263
;  PRIOR FILING DATE: 2000-05-31
;  PRIOR APPLICATION NUMBER: 60/208,831
;  PRIOR FILING DATE: 2000-06-02
;  PRIOR APPLICATION NUMBER: 60/209,451
;  PRIOR FILING DATE: 2000-06-05
;  PRIOR APPLICATION NUMBER: 60/210,060
;  PRIOR FILING DATE: 2000-06-07
;  PRIOR APPLICATION NUMBER: 60/219,507
;  PRIOR FILING DATE: 2000-07-20
;  PRIOR APPLICATION NUMBER: 60/221,337
;  PRIOR FILING DATE: 2000-07-26
;  PRIOR APPLICATION NUMBER: 60/221,927
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: 60/263,135
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: 60/263,688
;  PRIOR FILING DATE: 2001-01-24
;  PRIOR APPLICATION NUMBER: 60/263,694
;  PRIOR FILING DATE: 2001-01-24
;  NUMBER OF SEQ ID NOS: 155
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
;   LENGTH: 667
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-863-776-18

  Query Match             5.9%;  Score 190.5;  DB 10;  Length 667;
  Best Local Similarity   22.5%;  Pred. No. 3.1e-05;
  Matches  120;  Conservative   84;  Mismatches  179;  Indels  151;  Gaps   25;

Qy         28 GFRRTT------IAKREGAGDTEAD-----PSEQQPQQHNLSLRRSGRQPKRTERVEEFL 76
                |:| |       ::||      :: |     |||::  : :    ::  |     |:  :
Db        154 GLKRKTPALKVSVSKRARKASSDLDQASVSPSEEENSESSSESEKTSDQDFTPEK--KAA 211

Qy         77 TTVRRRG------KKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDSLGK 130
                 |||       ||  | : :  |:  |       |  |  |||   |  ||   |
Db        212 VRAPRRGPLGGRKKKKAPSASDSDSKADSDGAKPEPVAMARSASSSSSSSSSDSDVSVK 271

Qy        131 EHPASSE-------KAKGGEEEED------TSDSDSDGL-TLKELQNR---------LRR 167
              : |   :       | :| : : :      :||||||  :  : | : |           ||
Db        272 KPPRGRKPAEKPLPKPRGRKPKPERPPSSSSSDSDSDEVDRISEWKRRDEARRRELEARR 331

Qy        168 KREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSA-----EQDRPLCKQEPEA-SQG 221
              :||||    | ||  | :  |:|| | :|| ||    ||:     | | |: |:  :   :|
Db        332 RREQEEELRRLR-EQEKEEKERRRERAAERGSG--GSSGDELRED-DEPVKKRGRKGRGRG 388

Qy        222 PVSQSETDDIENQLEGKATQGNTEENPREAGK--PKPECEVYDPNALYCICRQPHNNRFM 279
              | | |:::  | :||          |||  |    ||:    :|      |:|
Db        389 PPSSSDSEP-EAELE-----------REAKKSAKKPQSSSTEP------ARKP------ 423

Qy        280 ICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQDSGCRS 339
                          | | | |                              |:    :
Db        424 --------------GQKEKRVR--------------------------PEEKQQAKP 440

Qy        340 VGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLK------IFQPVVEAPGAPK 393
              |  : |           :|  |   :  ::||    || ::||:       |    |::|   :
Db        441 VKVERT---------RKRSEGFSMDRKVEKKKEPSVEEKLQKLHSEIKFALKVDSPDVKR 491
```

```
Qy        394 CIGPGCSSVAQPDSVYCSNDCILKH--AAATMRFLSSGKEQKTKPKEKVKTKPEKFSLPK 451
              |:    ::: :  ::  ::  : |:    ||::     | :: |  : |  |  :
Db        492 CL----NALEELGTLQVTSQILQKNTDVVATLK-----KIRRYKANKDVMEKAAEVYTRL 542

Qy        452 CSVQVGIKISSVHK-RLASEKRENPVKKVMLASRSETSGKEAACESSTPSWASD 504
                |  :| || :| |   |  ::|   :|:   :  | :|:||  | :    ::|
Db        543 KSRVLGPKIEAVQKVDKAGMEKEKAEEKL---AGEELAGEEAPQEKAEDKPSTD 593
```

```
RESULT 10
US-10-106-698-5847
; Sequence 5847, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
;  APPLICANT: Ruben et al.
;  TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
;  FILE REFERENCE: PA005P1
;  CURRENT APPLICATION NUMBER: US/10/106,698
;  CURRENT FILING DATE:  2002-03-27
;  PRIOR APPLICATION NUMBER: PCT/US00/26524
;  PRIOR FILING DATE: 2000-09-28
;  PRIOR APPLICATION NUMBER: US 60/157,137
;  PRIOR FILING DATE: 1999-09-29
;  PRIOR APPLICATION NUMBER: US 60/163,280
;  PRIOR FILING DATE: 1999-11-03
;  NUMBER OF SEQ ID NOS: 8564
;  SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5847
;   LENGTH: 555
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-106-698-5847

  Query Match             5.8%;  Score 188.5;  DB 14;  Length 555;
  Best Local Similarity   21.6%;  Pred. No. 3.3e-05;
  Matches  116;  Conservative   86;  Mismatches  182;  Indels  153;  Gaps   24;

Qy         28 GFRRTT------IAKREGAGDTEAD-----PSEQQPQQHNLSLRRSGRQPKRTERVEEFL 76
                | :| |       ::||      :: |     |||::  : :    ::  |     |:  :
Db         38 GLKRKTPALKMSVSKRARKASSDLDQASVSPSEEENSESSSESEKTSDQDFTPEK--KAA 95

Qy         77 TTVRRRG------KKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDSLGK 130
                 |||       ||  | : :  |:  |       |  |  |||   |  ||   |
Db         96 VRAPRRGPLGGRKKKKAPSASDSDSKADSDGAKPEPVAMARSASSSSSSSSSDSDVSVK 155

Qy        131 EHPASSE-------KAKGGEEEED------TSDSDSDGL-TLKELQNR---------LRR 167
              : |   :       | :| : : :      :||||||  :  : | : |           ||
Db        156 KPPRGRKPAEKPLPKPRGRKPKPERPPSSSSSDSDSDEVDRISEWKRRDEARRRELEARR 215

Qy        168 KREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSA--------EQDRPLCKQEPEA- 218
              :||||    | ||  :   :::|||   |: |  : ||          | | |: |:  :
Db        216 RREQEEELRRLREQEKEEKERRRER--ADRGEAERGSGGSSGDELREDDEPVKKRGRKGR 273

Qy        219 SQGPVSQSETDDIENQLEGKATQGNTEENPREAGK--PKPECEVYDPNALYCICRQPHNN 276
              :|| | |:::  | :||             ||| |    ||:     :|       |:|
Db        274 GRGPPSSSDSEP-EAELE-----------REAKKSAKKPQSSSTEP------ARKP--- 311

Qy        277 RFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQDSG 336
                                | | | |                              |:
Db        312 -----------------GQKEKRVR--------------------------PEEKQQ 325

Qy        337 CRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLK------IFQPVVEAPG 390
              : |  : |           :|  |   :  ::||    || ::||:       |    |::|
Db        326 AKPVKVERT---------RKRSEGPSMDRKVEKKKEPSVEEKLQKLHSEIKFALKVDSPD 376

Qy        391 APKCIGPGCSSVAQPDSVYCSNDCILKH--AAATMRFLSSGKEQKTKPKEKVKTKPEKFS 448
              :|:    ::: :  ::  ::  : |:    ||::     | :: |  : |  |  :
Db        377 VKRCL----NALEELGTLQVTSQILQKNTDVVATLK-----KIRRYKANKDVMEKAAEVY 427

Qy        449 LPKCSVQVGIKISSVHK-RLASEKRENPVKKVMLASRSETSGKEAACESSTPSWASD 504
                |  :| || :| |   |  ::|   :|:   :  | :|:||  | :    ::|
Db        428 TRLKSRVLGPKIEAVQKVNKAGMEKEKAEEKL---AGEELAGEEAPQEKAEDKPSTD 481
```

```
RESULT 11
US-09-823-187-86
; Sequence 86, Application US/09823187
; Publication No. US20030096952A1
; GENERAL INFORMATION:
;  APPLICANT: Burgess, Catherine
;  APPLICANT:  Gusev, Vladimir Y
;  APPLICANT:  Liu, Xiaohong
;  APPLICANT:  Majumder, Kumud
;  APPLICANT:  Padigaru, Muralidhar
;  APPLICANT:  Patturajan, Meera
;  APPLICANT:  Shimkets, Richard A
;  APPLICANT:  Spaderna, Steven K
;  APPLICANT:  Spytek, Kimberly
;  APPLICANT:  Taupier, Raymond J
;  TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
;  FILE REFERENCE: 15966-745
;  CURRENT APPLICATION NUMBER: US/09/823,187
;  CURRENT FILING DATE:  2001-03-29
;  PRIOR APPLICATION NUMBER: 60/193,339
;  PRIOR FILING DATE: 2000-03-30
;  PRIOR APPLICATION NUMBER: 60/193,205
;  PRIOR FILING DATE: 2000-03-30
;  PRIOR APPLICATION NUMBER: 60/195,343
;  PRIOR FILING DATE: 2000-04-05
;  PRIOR APPLICATION NUMBER: 60/195,088
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: 60/195,005
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: 60/195,792
;  PRIOR FILING DATE: 2000-04-10
;  PRIOR APPLICATION NUMBER: 60/196,556
;  PRIOR FILING DATE: 2000-04-11
;  PRIOR APPLICATION NUMBER: 60/197,081
;  PRIOR FILING DATE: 2000-04-13
;  PRIOR APPLICATION NUMBER: 60/197,525
;  PRIOR FILING DATE: 2000-04-14
;  PRIOR APPLICATION NUMBER: 60/197,087
;  PRIOR FILING DATE: 2000-04-14
;  NUMBER OF SEQ ID NOS: 103
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 86
;   LENGTH: 670
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-823-187-86

  Query Match             5.8%;  Score 188.5;  DB 10;  Length 670;
  Best Local Similarity   21.6%;  Pred. No. 4.3e-05;
  Matches  116;  Conservative   86;  Mismatches  182;  Indels  153;  Gaps   24;

Qy          28 GFRRTT------IAKREGAGDTEAD-----PSEQQPQQHNLSLRRSGRQPKRTERVEEFL 76
               | :| |       ::||     :: |     |||::  : :    ::  |    |:  :
Db         154 GLKRKTPALKMSVSKRARKASSDLDQASVSPSEEENSESSSESEKTSDQDFTPEK--KAA 211

Qy          77 TTVRRRG------KKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSSEIRSGPVSDSLGK 130
                  |||       ||  | : :  |:  |          |   |  |||   |   ||   |
Db         212 VRAPRRGPLGGRKKKKAPSASDSDSKADSDGAKPEPVAMARSASSSSSSSSSSDSDVSVK 271

Qy         131 EHPASSE-------KAKGGEEEED------TSDSDSDGL-TLKELQNR---------LRR 167
               : |   :       | :| : : :      :||||| :  : | : |           ||
Db         272 KPPRGRKPAEKPLPKPRGRKPKPERPPSSSSSDSDSDEVDRISEWKRRDEARRRELEARR 331

Qy         168 KREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSA--------EQDRPLCKQEPEA- 218
               :||||   | ||  :  :::|||   |: |  : ||          | | |: |:  :
Db         332 RREQEEELRRLREQEKEKEKERRRER--ADRGEAERGSGGSSGDELRDDEPVKKRGRKGR 389

Qy         219 SQGPVSQSETDDIENQLEGKATQGNTEENPREAGK--PKPECEVYDPNALYCICRQPHNN 276
               :|| | |:::  | :||           ||| |   ||:    :|       |:|
Db         390 GRGPPSSSDSEP-EAELE-----------REAKKSAKKPQSSSTEP------ARKP--- 427

Qy         277 RFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQDSG 336
                                 | | | |                              |:
Db         428 -----------------GQKEKRVR----------------------------PEEKQQ 441

Qy         337 CRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLK------IFQPVVEAPG 390
                 : |  : |          :|  |   :  ::||   || ::||:       |   |::|
Db         442 AKPVKVERT---------RKRSEGFSMDRKVEKKKEPSVEEKLQKLHSEIKFALKVDSPD 492

Qy         391 APKCIGPGCSSVAQPDSVYCSNDCILKH--AAATMRFLSSGKEQKTKPKEKVKTKPEKFS 448
                  :|:    ::: :  ::  ::  : |:     ||::     | :: |  : |  |  :
Db         493 VKRCL----NALEELGTLQVTSQILQKNTDVVATLK-----KIRRYKANKDVMEKAAEVY 543

Qy         449 LPKCSVQVGIKISSVHK-RLASEKRENPVKKVMLASRSETSGKEAACESSTPSWASD 504
                   | :| || :| |   |  ::|   :|:   :  | :|:||  | :    ::|
Db         544 TRLKSRVLGPKIEAVQKVNKAGMEKEKAEEKL---AGEELAGEEAPQEKAEDKPSTD 597


RESULT 12
US-09-863-776-51
; Sequence 51, Application US/09863776
; Publication No. US20030198953A1
; GENERAL INFORMATION:
;  APPLICANT: Spytek, Kimberly A
;  APPLICANT:  Majumder, Kumud
;  APPLICANT:  Tchernev, Velizar T
;  APPLICANT:  Mishra, Vishnu
;  APPLICANT:  Padigaru, Muralidhara
;  APPLICANT:  Spaderna, Steven K
;  APPLICANT:  Shenoy, Suresh G
;  APPLICANT:  Rastelli, Luca
;  APPLICANT:  Li, Li
;  APPLICANT:  Taupier, Raymond J
;  APPLICANT:  Gangolli, Esha
;  TITLE OF INVENTION: No. US20030198953A1el Proteins and Nucleic Acids Encoding Same
;  FILE REFERENCE: 21402-020
;  CURRENT APPLICATION NUMBER: US/09/863,776
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: 09/540,763
;  PRIOR FILING DATE: 2000-03-30
;  PRIOR APPLICATION NUMBER: 60/206,679
;  PRIOR FILING DATE: 2000-05-24
;  PRIOR APPLICATION NUMBER: 60/206,688
;  PRIOR FILING DATE: 2000-05-24
;  PRIOR APPLICATION NUMBER: 60/206,829
;  PRIOR FILING DATE: 2000-05-24
;  PRIOR APPLICATION NUMBER: 60/207,748
;  PRIOR FILING DATE: 2000-05-30
;  PRIOR APPLICATION NUMBER: 60/207,798
;  PRIOR FILING DATE: 2000-05-30
;  PRIOR APPLICATION NUMBER: 60/208,263
;  PRIOR FILING DATE: 2000-05-31
;  PRIOR APPLICATION NUMBER: 60/208,831
;  PRIOR FILING DATE: 2000-06-02
;  PRIOR APPLICATION NUMBER: 60/209,451
;  PRIOR FILING DATE: 2000-06-05
;  PRIOR APPLICATION NUMBER: 60/210,060
;  PRIOR FILING DATE: 2000-06-07
;  PRIOR APPLICATION NUMBER: 60/219,507
;  PRIOR FILING DATE: 2000-07-20
;  PRIOR APPLICATION NUMBER: 60/221,337
;  PRIOR FILING DATE: 2000-07-26
;  PRIOR APPLICATION NUMBER: 60/221,927
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: 60/263,135
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: 60/263,688
;  PRIOR FILING DATE: 2001-01-24
;  PRIOR APPLICATION NUMBER: 60/263,694
;  PRIOR FILING DATE: 2001-01-24
;  NUMBER OF SEQ ID NOS: 155
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 51
;   LENGTH: 670
;   TYPE: PRT
```

```
;   ORGANISM: Homo sapiens
US-09-863-776-51

  Query Match             5.8%;  Score 188.5;  DB 10;  Length 670;
  Best Local Similarity   21.6%;  Pred. No. 4.3e-05;
  Matches  116;  Conservative   86;  Mismatches  182;  Indels  153;  Gaps   24;

Qy          28 GFRRTT------IAKREGAGDTEAD-----PSEQQPQQHNLSLRRSGRQPKRTERVEEPL 76
               | :| |       ::||     :: |     |||::  : :    ::  |    |:  :
Db         154 GLKRKTPALKMSVSKRARKASSDLDQASVSPSEEENSESSSESEKTSDQDFTPEK--KAA 211

Qy          77 TTVRRRG------KKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDSLGK 130
                 |||       ||  | : :  |:  |          |   |  |||   |   ||   |
Db         212 VRAPRRGPLGGRKKKKAPSASDSDSKADSDGAKPEPVAMARSASSSSSSSSSDSDVSVK 271

Qy         131 EHPASSE-------KAKGGEEEED------TSDSDSDGL-TLKELQNR---------LRR 167
               : |   :       | :| : : :      :|||||| :  : | : |          ||
Db         272 KPPRGRKPAEKPLPKPRGRKPKPERPPSSSSSDSDSDEVDRISEWKRRDEARRRELEARR 331

Qy         168 KREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSA--------EQDRPLCKQEPEA- 218
               :|||   | ||  :   :::|||   |: |  : ||         | | |: |:  :
Db         332 RREQEEELRRLREQEKEEKERRRER--ADRGEAERGSGGSSGDELRDDEPVKKRGRKGR 389

Qy         219 SQGPVSQSETDDIENQLEGKATQGNTEENPREAGK--PKPECEVYDPNALYCICRQPHNN 276
                :|| | |:::  | :||              |||  |   ||:    :|       |:|
Db         390 GRGPPSSSDSEP-EAELE-----------REAKKSAKKPQSSSTEP------ARKP--- 427

Qy         277 RFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQDSG 336
                                 | | | |                              |:
Db         428 -----------------GQKEKRVR----------------------------PEEKQQ 441

Qy         337 CRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLK------IFQPVVEAPG 390
                : |  : |         :|  |   :  ::||   || ::||:      |   |::|
Db         442 AKPVKVERT---------RKRSEGFSMDRKVEKKKEPSVEEKLQKLHSEIKFALKVDSPD 492

Qy         391 APKCIGPGCSSVAQPDSVYCSNDCILKH--AAATMRFLSSGKEQKTKPKEKVKTKPEKFS 448
                  :|:    ::: :  ::  ::  : |:     ||::     | :: |  : |  |  :
Db         493 VKRCL----NALEELGTLQVTSQILQKNTDVVATLK-----KIRRYKANKDVMEKAAEVY 543

Qy         449 LPKCSVQVGIKISSVHK-RLASEKRENPVKKVMLASRSETSGKEAACESSTPSWASD 504
                   | :| || :| |    |  ::|   :|:   :  | :|:||  | :    ::|
Db         544 TRLKSRVLGPKIEAVQKVNKAGMEKEKAEEKL---AGEELAGEEAPQEKAEDKPSTD 597


RESULT 13
US-09-946-374-308
; Sequence 308, Application US/09946374
; Publication No. US20030073129A1
; GENERAL INFORMATION:
;  APPLICANT: Baker, Kevin P.
;  APPLICANT:  Botstein, David
;  APPLICANT:  Desnoyers, Luc
;  APPLICANT:  Eaton, Dan L.
;  APPLICANT:  Ferrara, Napoleone
;  APPLICANT:  Fong, Sherman
;  APPLICANT:  Gao, Wei-Qiang
;  APPLICANT:  Goddard, Audrey
;  APPLICANT:  Godowski, Paul J.
;  APPLICANT:  Grimaldi, Christopher J.
;  APPLICANT:  Gurney, Austin L.
;  APPLICANT:  Hillan, Kenneth J.
;  APPLICANT:  Pan, James
;  APPLICANT:  Paoni, Nicholas F.
;  APPLICANT:  Roy, Margaret Ann
;  APPLICANT:  Smith, Victoria
;  APPLICANT:  Stewart, Timothy A.
;  APPLICANT:  Tumas, Daniel
;  APPLICANT:  Watanabe, Colin K.
;  APPLICANT:  Williams, P. Mickey
;  APPLICANT:  Wood, William I.
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: P2830P1C1
;  CURRENT APPLICATION NUMBER: US/09/946,374
;  CURRENT FILING DATE:  2001-09-04
;  PRIOR APPLICATION NUMBER: 60/098716
;  PRIOR FILING DATE: 1998-09-01
;  PRIOR APPLICATION NUMBER: 60/098723
;  PRIOR FILING DATE: 1998-09-01
;  PRIOR APPLICATION NUMBER: 60/098749
;  PRIOR FILING DATE: 1998-09-01
;  PRIOR APPLICATION NUMBER: 60/098750
;  PRIOR FILING DATE: 1998-09-01
;  PRIOR APPLICATION NUMBER: 60/098803
;  PRIOR FILING DATE: 1998-09-02
;  PRIOR APPLICATION NUMBER: 60/098821
;  PRIOR FILING DATE: 1998-09-02
;  PRIOR APPLICATION NUMBER: 60/098843
;  PRIOR FILING DATE: 1998-09-02
;  PRIOR APPLICATION NUMBER: 60/099536
;  PRIOR FILING DATE: 1998-09-09
;  PRIOR APPLICATION NUMBER: 60/099596
;  PRIOR FILING DATE: 1998-09-09
;  PRIOR APPLICATION NUMBER: 60/099598
;  PRIOR FILING DATE: 1998-09-09
;  PRIOR APPLICATION NUMBER: 60/099602
;  PRIOR FILING DATE: 1998-09-09
;  PRIOR APPLICATION NUMBER: 60/099642
;  PRIOR FILING DATE: 1998-09-09
;  PRIOR APPLICATION NUMBER: 60/099741
;  PRIOR FILING DATE: 1998-09-10
;  PRIOR APPLICATION NUMBER: 60/099754
;  PRIOR FILING DATE: 1998-09-10
;  PRIOR APPLICATION NUMBER: 60/099763
;  PRIOR FILING DATE: 1998-09-10
;  PRIOR APPLICATION NUMBER: 60/099792
;  PRIOR FILING DATE: 1998-09-10
;  PRIOR APPLICATION NUMBER: 60/099808
;  PRIOR FILING DATE: 1998-09-10
;  PRIOR APPLICATION NUMBER: 60/099812
;  PRIOR FILING DATE: 1998-09-10
;  PRIOR APPLICATION NUMBER: 60/099815
;  PRIOR FILING DATE: 1998-09-10
;  PRIOR APPLICATION NUMBER: 60/099816
;  PRIOR FILING DATE: 1998-09-10
;  PRIOR APPLICATION NUMBER: 60/100385
;  PRIOR FILING DATE: 1998-09-15
;  PRIOR APPLICATION NUMBER: 60/100388
;  PRIOR FILING DATE: 1998-09-15
;  PRIOR APPLICATION NUMBER: 60/100390
;  PRIOR FILING DATE: 1998-09-15
;  PRIOR APPLICATION NUMBER: 60/100584
;  PRIOR FILING DATE: 1998-09-16
;  PRIOR APPLICATION NUMBER: 60/100627
;  PRIOR FILING DATE: 1998-09-16
;  PRIOR APPLICATION NUMBER: 60/100661
;  PRIOR FILING DATE: 1998-09-16
;  PRIOR APPLICATION NUMBER: 60/100662
;  PRIOR FILING DATE: 1998-09-16
;  PRIOR APPLICATION NUMBER: 60/100664
;  PRIOR FILING DATE: 1998-09-16
;  PRIOR APPLICATION NUMBER: 60/100683
;  PRIOR FILING DATE: 1998-09-17
;  PRIOR APPLICATION NUMBER: 60/100684
;  PRIOR FILING DATE: 1998-09-17
;  PRIOR APPLICATION NUMBER: 60/100710
;  PRIOR FILING DATE: 1998-09-17
;  PRIOR APPLICATION NUMBER: 60/100711
;  PRIOR FILING DATE: 1998-09-17
;  PRIOR APPLICATION NUMBER: 60/100848
;  PRIOR FILING DATE: 1998-09-18
;  PRIOR APPLICATION NUMBER: 60/100849
;  PRIOR FILING DATE: 1998-09-18
;  PRIOR APPLICATION NUMBER: 60/100919
;  PRIOR FILING DATE: 1998-09-17
;  PRIOR APPLICATION NUMBER: 60/100930
```

; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/101014
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/101068
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/101071
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/101279
; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: 60/101471
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101472
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101474
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101475
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101476
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101477
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101479
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101738
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101741
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101743
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101915
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101916
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/102207
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102240
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102307
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102330
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102331
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102484
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/102487
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/102570
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/102571
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/102684
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/102687
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/102965
; PRIOR FILING DATE: 1998-10-02
; PRIOR APPLICATION NUMBER: 60/103258
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: 60/103314
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103315
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103328
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103395
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103396
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103401
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103449
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: 60/103633
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/103678
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/103679
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/103711
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/104257
; PRIOR FILING DATE: 1998-10-14
; PRIOR APPLICATION NUMBER: 60/104987
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105000
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105002
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105104
; PRIOR FILING DATE: 1998-10-21
; PRIOR APPLICATION NUMBER: 60/105169
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/105266
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/105693
; PRIOR FILING DATE: 1998-10-26
; PRIOR APPLICATION NUMBER: 60/105694
; PRIOR FILING DATE: 1998-10-26
; PRIOR APPLICATION NUMBER: 60/105807

Query Match 5.8%; Score 188.5; DB 10; Length 671;
Best Local Similarity 21.6%; Pred. No. 4.3e-05;
Matches 116; Conservative 86; Mismatches 182; Indels 153; Gaps 24;

```
Qy          28 GFRRTT------IAKREGAGDTEAD-----PSEQQPQQHNLSLRRSGRQPKRTERVEEFL 76
               | :| |       ::||      :: |     |||::  : :    ::  |    |:  :
Db         154 GLKRKTPALKMSVSKRARKASSDLDQASVSPSEEENSESSSESEKTSDQDFTPEK--KAA 211

Qy          77 TTVRRRG------KKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDSLGK 130
                 |||       ||  | : :  |: |          |   |  |||   |   ||   |
Db         212 VRAPRRGPLGGRKKKKAPSASDSDSKADSDGAKPEPVAMARSASSSSSSSSSDSDVSVK 271

Qy         131 EHPASSE-------KAKGGEEEED------TSDSDSDGL-TLKELQNR---------LRR 167
               : |   :       | :| : : :      :||||||  :  : | : |           ||
Db         272 KPPRGRKPAEKPLPKPRGRKPKPERPPSSSSSDSDSDEVDRISEWKRRDEARRRELEARR 331

Qy         168 KREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSA--------EQDRPLCKQEPEA- 218
               :||||   | ||  :  :::|||   |: |  : ||          | | |: |:  :
Db         332 RREQEEELRRLREQEKEEKERRRER--ADRGEAERGSGGSSGDELRREDDEPVKKRGRKGR 389

Qy         219 SQGPVSQSETDDIENQLEGKATQGNTEENPREAGK--PKPECEVYDPNALYCICRQPHNN 276
               :|| | |:::  | :||                |||  |   ||:    :|       |:|
Db         390 GRGPPSSSDSEP-EAELE------------RRAKKSAKKPQSSSTEP------ARKP--- 427

Qy         277 RFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQDSG 336
                                 | | | |                          |:
Db         428 -----------------GQKEKRVR----------------------------PEEKQQ 441

Qy         337 CRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLK------IFQPVVEAPG 390
               : |  : |           :|  |   :  ::||   || ::||:       |   |::|
Db         442 AKPVKVERT---------RKRSEGFSMDRKVEKKKEPSVEEKLQKLHSEIKFALKVDSPD 492

Qy         391 APKCIGPGCSSVAQPDSVYCSNDCILKH--AAATMRFLSSGKEQKTKPKEKVKTKPEKFS 448
                 :|:    ::: :  ::  ::  : |:    ||::     | :: |  : | |  :
Db         493 VKRCL----NALEELGTLQVTSQILQKNTDVVATLK-----KIRRYKANKDVMEKAAEVY 543

Qy         449 LPKCSVQVGIKISSVHK-RLASEKRENPVKKVMLASRSETSGKEAACESSTPSWASD 504
                   |  :| || :| |   |  ::|   :|:   :  | :|:||  | :    ::|
Db         544 TRLKSRVLGPKIEAVQKVNKAGMEKEKAEEKL---AGEELAGEEAPQEKAEDKPSTD 597
```

RESULT 14
US-09-823-187-85
; Sequence 85, Application US/09823187
; Publication No. US20030096952A1

```
; GENERAL INFORMATION:
;  APPLICANT: Burgess, Catherine
;  APPLICANT:  Gusev, Vladimir Y
;  APPLICANT:  Liu, Xiaohong
;  APPLICANT:  Majumder, Kumud
;  APPLICANT:  Padigaru, Muralidhar
;  APPLICANT:  Patturajan, Meera
;  APPLICANT:  Shimkets, Richard A
;  APPLICANT:  Spaderna, Steven K
;  APPLICANT:  Spytek, Kimberly
;  APPLICANT:  Taupier, Raymond J
;  TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
;  FILE REFERENCE: 15966-745
;  CURRENT APPLICATION NUMBER: US/09/823,187
;  CURRENT FILING DATE:  2001-03-29
;  PRIOR APPLICATION NUMBER: 60/193,339
;  PRIOR FILING DATE: 2000-03-30
;  PRIOR APPLICATION NUMBER: 60/193,205
;  PRIOR FILING DATE: 2000-03-30
;  PRIOR APPLICATION NUMBER: 60/195,343
;  PRIOR FILING DATE: 2000-04-05
;  PRIOR APPLICATION NUMBER: 60/195,088
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: 60/195,005
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: 60/195,792
;  PRIOR FILING DATE: 2000-04-10
;  PRIOR APPLICATION NUMBER: 60/196,556
;  PRIOR FILING DATE: 2000-04-11
;  PRIOR APPLICATION NUMBER: 60/197,081
;  PRIOR FILING DATE: 2000-04-13
;  PRIOR APPLICATION NUMBER: 60/197,525
;  PRIOR FILING DATE: 2000-04-14
;  PRIOR APPLICATION NUMBER: 60/197,087
;  PRIOR FILING DATE: 2000-04-14
;  NUMBER OF SEQ ID NOS: 103
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 85
;   LENGTH: 671
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-823-187-85

  Query Match             5.8%;  Score 188.5;  DB 10;  Length 671;
  Best Local Similarity   21.6%;  Pred. No. 4.3e-05;
  Matches  116;  Conservative   86;  Mismatches  182;  Indels  153;  Gaps   24;

Qy         28 GFRRTT------IAKREGAGDTEAD-----PSEQQPQQHNLSLRRSGRQPKRTERVEEFL 76
              | :| |      ::||     :: |     |||::  : :    ::  |    |:  :
Db        154 GLKRKTPALKMSVSKRARKASSDLDQASVSPSEEENSESSSESEKTSDQDFTPEK--KAA 211

Qy         77 TTVRRRG------KKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDSLGK 130
                 |||      ||  | : :  |:  |          |   |  |||   |   ||   |
Db        212 VRAPRRGPLGGRKKKKAPSASDSDSKADSDGAKPEPVAMARSASSSSSSSSSSDSDVSVK 271

Qy        131 EHPASSE-------KAKGGEEEED------TSDSDSDGL-TLKELQNR---------LRR 167
              : |   :       | :| : : :      :||||| :  : | : |          ||
Db        272 KPPRGRKPAEKPLPKPRGRKPKPERPPSSSSSDSDSDEVDRISEWKRRDEARRRELEARR 331

Qy        168 KREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSA--------EQDRPLCKQEPEA- 218
              :||||    | ||  :   :::|||   |: |  : ||          | | |: |:  :
Db        332 RREQEEELRRLREQEKEEKERRRER--ADRGEAERGSGGSSGDELREDDEPVKKRGRKGR 389

Qy        219 SQGPVSQSETDDIENQLEGKATQGNTEENPREAGK--PKPECEVYDPNALYCICRQPHNN 276
              :|| | |:::  | :||           |||| |   ||:    :|       |:|
Db        390 GRGPPSSSDSEP-EAELE-----------REAKKSAKKPQSSSTEP------ARKP--- 427

Qy        277 RFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQDSG 336
                          | | | |                                    |:
Db        428 -----------------GQKEKRVR---------------------------PEEKQQ 441

Qy        337 CRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLK------IFQPVVEAPG 390
              : |  : |         :|  |   :  ::||   || ::||:       |   |::|
```

```
Db        442 AKPVKVERT---------RKRSEGPSMDRKVEKKKEPSVEEKLQKLHSEIKPALKVDSPD 492

Qy        391 APKCIGPGCSSVAQPDSVYCSNDCILKH--AAATMRFLSSGKEQKTKPKEKVKTKPEKFS 448
                :|:    ::: :  ::  ::  : |:    ||::     | :: |  : | | :
Db        493 VKRCL----NALEELGTLQVTSQILQKNTDVVATLK-----KIRRYKANKDVMEKAAEVY 543

Qy        449 LPKCSVQVGIKISSVHK-RLASEKRENPVKKVMLASRSETSGKEAACESSTPSWASD 504
                  |  :| || :| |   |  ::|   :|:   :  | :|:||  | :    ::|
Db        544 TRLKSRVLGPKIEAVQKVNKAGMEKEKAEEKL---AGEELAGEEAPQEKAEDKPSTD 597


RESULT 15
US-09-863-776-20
; Sequence 20, Application US/09863776
; Publication No. US20030198953A1
; GENERAL INFORMATION:
;  APPLICANT: Spytek, Kimberly A
;  APPLICANT:  Majumder, Kumud
;  APPLICANT:  Tchernev, Velizar T
;  APPLICANT:  Mishra, Vishnu
;  APPLICANT:  Padigaru, Muralidhara
;  APPLICANT:  Spaderna, Steven K
;  APPLICANT:  Shenoy, Suresh G
;  APPLICANT:  Rastelli, Luca
;  APPLICANT:  Li, Li
;  APPLICANT:  Taupier, Raymond J
;  APPLICANT:  Gangolli, Esha
;  TITLE OF INVENTION: No. US20030198953A1el Proteins and Nucleic Acids Encoding Same
;  FILE REFERENCE: 21402-020
;  CURRENT APPLICATION NUMBER: US/09/863,776
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: 09/540,763
;  PRIOR FILING DATE: 2000-03-30
;  PRIOR APPLICATION NUMBER: 60/206,679
;  PRIOR FILING DATE: 2000-05-24
;  PRIOR APPLICATION NUMBER: 60/206,688
;  PRIOR FILING DATE: 2000-05-24
;  PRIOR APPLICATION NUMBER: 60/206,829
;  PRIOR FILING DATE: 2000-05-24
;  PRIOR APPLICATION NUMBER: 60/207,748
;  PRIOR FILING DATE: 2000-05-30
;  PRIOR APPLICATION NUMBER: 60/207,798
;  PRIOR FILING DATE: 2000-05-30
;  PRIOR APPLICATION NUMBER: 60/208,263
;  PRIOR FILING DATE: 2000-05-31
;  PRIOR APPLICATION NUMBER: 60/208,831
;  PRIOR FILING DATE: 2000-06-02
;  PRIOR APPLICATION NUMBER: 60/209,451
;  PRIOR FILING DATE: 2000-06-05
;  PRIOR APPLICATION NUMBER: 60/210,060
;  PRIOR FILING DATE: 2000-06-07
;  PRIOR APPLICATION NUMBER: 60/219,507
;  PRIOR FILING DATE: 2000-07-20
;  PRIOR APPLICATION NUMBER: 60/221,337
;  PRIOR FILING DATE: 2000-07-26
;  PRIOR APPLICATION NUMBER: 60/221,927
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: 60/263,135
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: 60/263,688
;  PRIOR FILING DATE: 2001-01-24
;  PRIOR APPLICATION NUMBER: 60/263,694
;  PRIOR FILING DATE: 2001-01-24
;  NUMBER OF SEQ ID NOS: 155
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
;   LENGTH: 671
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-863-776-20

  Query Match             5.8%;  Score 188.5;  DB 10;  Length 671;
  Best Local Similarity   21.6%;  Pred. No. 4.3e-05;
```

Matches 116; Conservative 86; Mismatches 182; Indels 153; Gaps 24;

```
Qy         28 GFRRTT------IAKREGAGDTEAD-----PSEQQPQQHNLSLRRSGRQPKRTERVEEFL 76
Db        154 GLKRKTPALKMSVSKRARKASSDLDQASVSPSEEENSESSSESEKTSDQDFTPEK--KAA 211

Qy         77 TTVRRRG------KKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSSEIRSGPVSDSLGK 130
Db        212 VRAPRRGPLGGRKKKKAPSASDSDSKADSDGAKPEPVAMARSASSSSSSSSSSDSDVSVK 271

Qy        131 EHPASSE-------KAKGGEEEED------TSDSDSDGL-TLKELQNR---------LRR 167
Db        272 KPPRGRKPAEKPLPKPRGRKPKPERPPSSSSSDSDSDEVDRISEWKRRDEARRRELEARR 331

Qy        168 KREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSA--------EQDRPLCKQEPEA- 218
Db        332 RREQEEELRRLREQEKEEKERRRER--ADRGEAERGSGGSSGDELREDDEPVKKRGRKGR 389

Qy        219 SQGPVSQSETDDIENQLEGKATQGNTEENPREAGK--PKPECEVYDPNALYCICRQPHNN 276
Db        390 GRGPPSSSDSEP-EAELE------------REAKKSAKKPQSSSTEP------ARKP--- 427

Qy        277 RFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQDSG 336
Db        428 -----------------GQKEKRVR--------------------------PEEKQQ 441

Qy        337 CRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLK------IFQPVVEAPG 390
Db        442 AKPVKVERT---------RKRSEGFSMDRKVEKKKEPSVEEKLQKLHSEIKFALKVDSPD 492

Qy        391 APKCIGPGCSSVAQPDSVYCSNDCILKH--AAATMRFLSSGKEQKTKPKEKVKTKPEKPS 448
Db        493 VKRCL----NALEELGTLQVTSQILQKNTDVVATLK-----KIRRYKANKDVMEKAAEVY 543

Qy        449 LPKCSVQVGIKISSVHK-RLASEKRENPVKKVMLASRSETSGKEAACESSTPSWASD 504
Db        544 TRLKSRVLGPKIEAVQKVNKAGMEKEKAEEKL---AGEELAGEEAPQEKAEDKPSTD 597
```

Search completed: April 20, 2004, 21:25:48
Job time : 51 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 20, 2004, 21:17:09 ; Search time 23 Seconds
(without alignments)
1378.189 Million cell updates/sec

Title: US-09-787-016A-4
Perfect score: 3228
Sequence: 1 MDDKGHLSNEEAPKAIKPTS.....LKSCVGLMLEAISYFSFRPW 614

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	197.5	6.1	1805	1	US-07-853-913-2	Sequence 2, Appli
2	166	5.1	688	3	US-09-141-047-8	Sequence 8, Appli
3	165.5	5.1	1780	1	US-08-769-309A-5	Sequence 5, Appli
4	165.5	5.1	1780	3	US-08-994-570-5	Sequence 5, Appli
5	164	5.1	238	4	US-09-257-179-80	Sequence 80, Appl
6	153.5	4.8	1848	3	US-08-296-791-6	Sequence 6, Appli
7	153.5	4.8	1848	4	US-09-839-996-6	Sequence 6, Appli
8	153.5	4.8	1848	4	US-10-080-505-6	Sequence 6, Appli
9	153.5	4.8	1848	5	PCT-US95-10661A-6	Sequence 6, Appli
10	151.5	4.7	532	1	US-08-285-440-5	Sequence 5, Appli
11	151.5	4.7	532	1	US-08-630-349-5	Sequence 5, Appli
12	149.5	4.6	2468	4	US-09-976-594-726	Sequence 726, App
13	148	4.6	657	3	US-08-893-852A-3	Sequence 3, Appli
14	148	4.6	657	3	US-08-821-818-3	Sequence 3, Appli
15	148	4.6	657	4	US-09-052-753B-3	Sequence 3, Appli
16	144.5	4.5	1088	3	US-09-130-242-2	Sequence 2, Appli
17	142	4.4	1596	4	US-08-978-277A-4	Sequence 4, Appli
18	141.5	4.4	1739	4	US-09-976-594-76	Sequence 76, Appl
19	141	4.4	8991	4	US-08-714-741-32	Sequence 32, Appl
20	140.5	4.4	558	1	US-08-285-440-6	Sequence 6, Appli
21	140.5	4.4	558	1	US-08-630-349-6	Sequence 6, Appli
22	138.5	4.3	2289	3	US-09-051-019-2	Sequence 2, Appli
23	135.5	4.2	1233	4	US-09-645-456A-35	Sequence 35, Appl
24	135.5	4.2	1233	4	US-09-425-324A-35	Sequence 35, Appl
25	135.5	4.2	1233	4	US-09-645-791-35	Sequence 35, Appl
26	134.5	4.2	1231	4	US-09-595-684B-23	Sequence 23, Appl
27	134.5	4.2	1297	4	US-09-688-188B-14	Sequence 14, Appl
28	134.5	4.2	1297	4	US-09-291-417D-14	Sequence 14, Appl
29	134.5	4.2	1332	4	US-09-645-456A-9	Sequence 9, Appli
30	134.5	4.2	1332	4	US-09-425-324A-9	Sequence 9, Appli
31	134.5	4.2	1332	4	US-09-645-791-9	Sequence 9, Appli
32	134	4.2	1075	4	US-09-252-991A-18387	Sequence 18387, A
33	134	4.2	3256	4	US-09-919-172-98	Sequence 98, Appl
34	134	4.2	3256	4	US-09-976-594-22	Sequence 22, Appl
35	132.5	4.1	491	1	US-07-903-103-2	Sequence 2, Appli
36	132.5	4.1	491	1	US-08-044-619A-2	Sequence 2, Appli
37	132.5	4.1	491	1	US-08-283-911-2	Sequence 2, Appli
38	132.5	4.1	491	1	US-08-245-500A-3	Sequence 3, Appli
39	132.5	4.1	491	1	US-08-390-546-3	Sequence 3, Appli
40	132.5	4.1	491	1	US-08-390-479A-3	Sequence 3, Appli
41	132.5	4.1	491	1	US-08-557-393-3	Sequence 3, Appli
42	132.5	4.1	491	1	US-08-390-516C-3	Sequence 3, Appli
43	132.5	4.1	491	1	US-08-390-517A-3	Sequence 3, Appli
44	132.5	4.1	491	1	US-08-390-515A-3	Sequence 3, Appli
45	132.5	4.1	491	2	US-08-801-718-3	Sequence 3, Appli

ALIGNMENTS

```
RESULT 1
US-07-853-913-2
; Sequence 2, Application US/07853913
; Patent No. 5338839
; GENERAL INFORMATION:
;   APPLICANT: McKay, Ronald D.G.
;   APPLICANT: Lendahl, Urban
;   TITLE OF INVENTION: Nestin Expression As An Indicator of
;   TITLE OF INVENTION: Neuroepithelial Tumors
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
;     STREET: Two Militia Drive
;     CITY: Lexington
;     STATE: Massachusetts
;     COUNTRY: U.S.A.
;     ZIP: 02173
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/07/853,913
;     FILING DATE: 19920319
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/660,412
;     FILING DATE: 22-FEB-1991
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/603,803
;     FILING DATE: 25-OCT-1990
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/201,762
;     FILING DATE: 02-JUN-1988
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/180,548
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Granahan, Patricia
;     REGISTRATION NUMBER: 32,227
;     REFERENCE/DOCKET NUMBER: MIT-4641AAAA
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 617-861-6240
;     TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1805 amino acids
;     TYPE: AMINO ACID
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
```

US-07-853-913-2

Query Match 6.1%; Score 197.5; DB 1; Length 1805;
Best Local Similarity 21.2%; Pred. No. 8.3e-08;
Matches 143; Conservative 99; Mismatches 223; Indels 211; Gaps 35;

```
Qy          2 DDKG---HLSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPSEQQPQQHNLS 58
              :|:|     |  :||   :|  :   :||          :|  |||     | | |  :
Db        486 EDEGQIWELVEKEADIEVKVENSSAQKT----------QESGLDTEETQDSQGPLQKE-T 534

Qy         59 LRRSGRQPKRTERVEEFLTTVRRRGKKNVPVSLEDS----SEPTSSTVTDVETASEGSVE 114
              |:  | :|  : ::: : |     ||:|   | |          |       ::: | :||
Db        535 LKALGEEPLMSLKIQNYETA----GKENCNSSTEGHLGTLEGPEKEKQIPLKSLEEKNVE 590

Qy        115 SSSEIRSG-PV-SDSLGKEHPASSEK----AKG--------GEEEEDTSDSDSDG----- 155
              |   : :| || |: ||||   : ::     ||         |:| ::      :|
Db        591 SEKTLENGVPVLSELLGKEDTRTEDQELMSPKGTLKRFSSLGKESQEVVRPSKEGNLESW 650

Qy        156 ----------------------LTLKELQN------------RLRRKREQEP------ 173
                                   |  || |:            |  :|  |||
Db        651 TAFKEESQHPLGFPGAEDQMLERLVEKEDQSFPRSPEEEDQEACRPLQKENQEPLGYEEA 710

Qy        174 ----VERSL-RGSQNRLRKKRREEDSAETGSVQIGSAEQDRPLCKQEPEASQGPVSQSET 228
                  :|| : : ||  ||    ||:  | |           | | |:    :| |:   |
Db        711 EGQILERLIEKESQESLRSP--EEEDQEAG----------RSLQKE----NQEPLGYEEA 754

Qy        229 DD--IENQLEGKATQG--NTEENPREAGKP-----KPECEVYDPNALYCICRQPHNNRFM 279
              :|  :|  :| :: :   : ||| |  |||      :      : |    :  :  | | :
Db        755 EDQMLERLIEKESQESLKSPEENQR-IGKPLERENQKSLRYLEENQETFVPLESRNQRPL 813

Qy        280 ICCDRCEE----------------------------WFHGDCVGIS-------EARGRL 303
              :  ||                             ||:  |          :: | |
Db        814 RSLEVEEEEQRIVKPLEKVSQDSLGSLAEENVQPLRYLEEDDCINKSLLEDKTHKSLGSL 873

Qy        304 LERNGEDYICPNCTILQV----------------QDETNGSATNEQDSGCRSVGADGTD 346
              :|||:  | |  :  ||                ::    | ::|::          :| :
Db        874 EDRNGDSIIIPQESETQVSLRPPEEEDQRIVNHLEKESQEPSRSSEEEEQVMERSLEGEN 933

Qy        347 CTSIGTVEQKSGEDQGIKGRIEKAANPSGK----KKLKIFQPVVEAPGAPKCIGPGCSSV 402
              |: :||:   ||| :: ::|| :  |||     :  : | |:        |        |:
Db        934 HESLSSVEK---EDQMVESQLEKESQDSGKSLEDESQETFGPL------EKENAESLRSL 984

Qy        403 AQPDSVYCSNDCILKHAAATMRFLSSGKEQ-KTKPKEKVKTKPEKFSLPKCSVQVGIKIS 461
              | |          : :     |:|  : | ||    | |||    ||   |||     :|
Db        985 AGQDQ---EEQKLEQETQQTLR--AVGNEQMAVSPPEKV--DPE---LPK---PLG-NDQ 1030

Qy        462 SVHKRLASEKREN----------PVKKV------MLASRSETSGKEAACESSTPSWASDH 505
              : : |  | :|:           || :      | :  |   :  : ::  | |:::
Db       1031 EIARSLGKENQESLVSLKEKGIETVKSLETEIIEPLETAEEDLERRKSIDTQEPLWSTEV 1090

Qy        506 NYNAVKPEKPEKPTAL 521
                |:| : | | :|
Db       1091 ARETVEPPEDEPPGSL 1106
```

RESULT 2
US-09-141-047-8
; Sequence 8, Application US/09141047A
; Patent No. 6043085
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Ehrlichia canis 120-kDa Immunodominant Antigenic
; TITLE OF INVENTION: Protein Gene
; FILE REFERENCE: D6143
; CURRENT APPLICATION NUMBER: US/09/141,047A
; CURRENT FILING DATE: 1998-08-27
; NUMBER OF SEQ ID NOS: 15
; SEQ ID NO 8
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of 120 kDa immunoreactive
; OTHER INFORMATION: protein.
US-09-141-047-8

Query Match 5.1%; Score 166; DB 3; Length 688;
Best Local Similarity 21.4%; Pred. No. 1.1e-05;
Matches 128; Conservative 84; Mismatches 211; Indels 176; Gaps 30;

```
Qy          8 SNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPSEQQPQQHNLSLRRSGRQPK 67
              |: |  | :  ||||          |   | |     :  |:     :|: |    | :
Db         90 SSSEVGKKVSETSKE-------ESTPEVKAE-----DLQPAVDGSIEHSSS--EVGEKVS 135

Qy         68 RTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTV-----------------TDVETAS 109
              :| : ||     |:    :::  :::|| | :|| |                      |:: |
Db        136 KTSK-EESTPEVK---AEDLQPAVDDSVEHSSSEVGEKVSETSKEENTPEVKAEDLQPAV 191

Qy        110 EGSVE-SSSEI----------RSGP----------VSDSLGKEHPAS--SEKAKGGEEEE 146
              :||:| ||||:          | |          | ||:  || :|   ||      :||
Db        192 DGSIEHSSSEVGEKVSKTSKEESTPEVKAEDLQPAVDDSV--EHSSSEVGEKVSETSKEE 249

Qy        147 DTSDSDSDGLTLKELQNRLRRKREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSAE 206
              :| :  :: |               | |: |:  | : : ||   | | |  : ::  ||
Db        250 NTPEVKAEDL-------------QPAVDGSVEHSSSEVGKK-VSETSKEESTPEV-KAE 293

Qy        207 QDRPLCKQEPEASQGPVSQ--SETDDIENQLEGK------ATQGNTEENPREAG------ 252
              :|       | |   | :  |||   ||  | :       | |: | :  | |
Db        294 DLQPAVDDSVEHSSSEVGEKVSETSKEENTPEVRAEDLQPAVDGSVEHSSSEVGEKVSET 353

Qy        253 ---KPKPECEVYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGE 309
                 :  || :  |          ||          |   |       |   |: :    :
Db        354 SKEESTPEVKAED--------LQP-------AVDSSIE--HS-----SSEVGKKVSETSK 391

Qy        310 DYICPNCTILQVQDETNGSATNEQDSGCRSVGADGTDCTSIGTVEQKSGED--QGIKGRI 367
              :   |       :|  :||  :         ||   :: :       :   ||    : | :
Db        392 EESTPEVKAEDLQPAVDGSVEHSSS----EVGEKVSETSKEENTPEVKAEDLQPAVDGSV 447

Qy        368 EKAANPSGKK-------------KLKIFQPVVEAPGAPKCIGPGCSSVAQPDSVYCSNDC 414
              | :::  |:|                | :  || |                        |
Db        448 EHSSSEVGEKVSETSKEENTPEVKAEDLQPAV-------------------------DG 481

Qy        415 ILKHAAATM--RPLSSGKEQKTKPKEKVKTKPE-KFSLPKCSVQVGIKISSVHK-----R 466
              ::|::: :  :   : ||: |  :    :|    |:   | :|| |:|  .|
Db        482 SVEHSSSEVGEKVSETSKEESTPEVKAEDLQPAVDGSVEHSSSEVGEKVSETSKEESTPE 541

Qy        467 LASEKRENPVKKVMLASRSETSGK--EAACESSTPSWASDHNYNAVKPEKPEKPTALSP 523
              : :|  :  |  :  | ||   |  | : | |||   ::     |:|     |  |:|
Db        542 VKAEDLQPAVDGSVEHSSSEVGEKVSETSKEESTPEVKAE-----VQPVADGNPVPLNP 595
```

RESULT 3
US-08-769-309A-5
; Sequence 5, Application US/08769309A
; Patent No. 5741890
; GENERAL INFORMATION:
; APPLICANT: Scott, John D.,
; APPLICANT: Nauert, Brian J.,
; APPLICANT: Klauck, Theresa M.
; TITLE OF INVENTION: Protein Binding Domains of Gravin
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/769,309A

```
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5741890and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27866/33451
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-474-6300
;      TELEFAX:  312-474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1780 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-769-309A-5

  Query Match             5.1%;  Score 165.5;  DB 1;  Length 1780;
  Best Local Similarity   20.4%;  Pred. No. 4.8e-05;
  Matches  133;  Conservative   81;  Mismatches  201;  Indels  237;  Gaps   32;

Qy          3 DKGHLSN---EEAPKA-------IKPTSKE-----FRKTW---GFRRT------------ 32
Db        136 DDGQEENRNIEQIPSSESNLEELTQPTESQANDIGFKKVFKFVGFKFTVKKDKTEKPDTV 195

Qy         33 ---TIAKREGAGDTEA----DPS---------EQQPQQHNLSLRRSGRQPKRTERVEEFL 76
Db        196 QLLTVKKDEGEGAAGAGDHQDPSLGAGEAASKESEPKQ-------------STEKPEETL 242

Qy         77 TTVRRRGKKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDSLGK------ 130
Db        243 KREQSHAEISPPA---ESGQAVEECKEEGEEKQEKEPSKSAESPTSPVTSETGSTFKKFF 299

Qy        131 -------EHPASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLRRKREQEPVERSLRGSQN 183
Db        300 TQGWAGWRKKTSFRKPK--EDEVEASE---------------KKKEQEP---------- 331

Qy        184 RLRKKRREED-SAETGSVQIGSAEQDRPLCKQEP-EASQGPVSQSETDDI----ENQLEG 237
Db        332 --EKVDTEEDGKAEVASEKLTASEQAHP---QEPAESAHEPRLSAEYEKVELPSEEQVSG 386

Qy        238 KATQGNTEENPREAGKPKP-ECEVYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGI 296
Db        387 --SQGPSEE------KPAPLATEVFDEKI------EVHQEEVVA-----------EVHV 420

Qy        297 SEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQDSGCRSVGADGTDCTSIGTVEQK 356
Db        421 STVEERTEEQKTE-----------VEETAGSVPAEELVGMDAEPQEAEPAKELVKLKET 468

Qy        357 --SGED--QGI------------KGRIEKAANPSGKKKLKI----FQPVVEAPGAPKCI 395
Db        469 CVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERMKVQGSPLKKLFTSTGLKKLS 528

Qy        396 G---------------------------------PGCSSVAQPDSVYCSNDCILKHAAA 421
Db        529 GKKQKGKRGGGDEESGEHTQVPADSPDSQEEQKGESSASSPEEPEEITC-----LEKGLA 583

Qy        422 TMRFLSSGKEQKTKPKEKVK------TKPEKPSLPKCSVQVGIKISSVHKRLASEKRENP 475
Db        584 EVQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRV----------RRPSESDKEDE 633

Qy        476 VKKVMLASRSETSGKEAACESSTPSWASDHNYNAVKPEKPEKPTALSPTLLS 527
Db        634 LDKVKSATLSST--------ESTASEMQEEMKGSVEEPKPEEPKRKVDTSVS 677


RESULT 4
US-08-994-570-5
; Sequence 5, Application US/08994570
; Patent No. 6090929
;  GENERAL INFORMATION:
;    APPLICANT:  Scott, John D.,
;    APPLICANT:  Nauert, Brian J.,
;    APPLICANT:  Klauck, Theresa M.
;    TITLE OF INVENTION:  Protein Binding Domains of Gravin
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower/233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/994,570
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6090929and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27866/33451
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-474-6300
;      TELEFAX:  312-474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1780 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-994-570-5

  Query Match             5.1%;  Score 165.5;  DB 3;  Length 1780;
  Best Local Similarity   20.4%;  Pred. No. 4.8e-05;
  Matches  133;  Conservative   81;  Mismatches  201;  Indels  237;  Gaps   32;

Qy          3 DKGHLSN---EEAPKA-------IKPTSKE-----FRKTW---GFRRT------------ 32
Db        136 DDGQEENRNIEQIPSSESNLEELTQPTESQANDIGFKKVFKFVGFKFTVKKDKTEKPDTV 195

Qy         33 ---TIAKREGAGDTEA----DPS---------EQQPQQHNLSLRRSGRQPKRTERVEEFL 76
Db        196 QLLTVKKDEGEGAAGAGDHQDPSLGAGEAASKESEPKQ-------------STEKPEETL 242

Qy         77 TTVRRRGKKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDSLGK------ 130
Db        243 KREQSHAEISPPA---ESGQAVEECKEEGEEKQEKEPSKSAESPTSPVTSETGSTFKKFF 299

Qy        131 -------EHPASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLRRKREQEPVERSLRGSQN 183
Db        300 TQGWAGWRKKTSFRKPK--EDEVEASE---------------KKKEQEP---------- 331

Qy        184 RLRKKRREED-SAETGSVQIGSAEQDRPLCKQEP-EASQGPVSQSETDDI----ENQLEG 237
Db        332 --EKVDTEEDGKAEVASEKLTASEQAHP---QEPAESAHEPRLSAEYEKVELPSEEQVSG 386

Qy        238 KATQGNTEENPREAGKPKP-ECEVYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGI 296
Db        387 --SQGPSEE------KPAPLATEVFDEKI------EVHQEEVVA-----------EVHV 420

Qy        297 SEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQDSGCRSVGADGTDCTSIGTVEQK 356
Db        421 STVEERTEEQKTE-----------VEETAGSVPAEELVGMDAEPQEAEPAKELVKLKET 468

Qy        357 --SGED--QGI------------KGRIEKAANPSGKKKLKI----FQPVVEAPGAPKCI 395
Db        469 CVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERMKVQGSPLKKLFTSTGLKKLS 528

Qy        396 G---------------------------------PGCSSVAQPDSVYCSNDCILKHAAA 421
```

```
Db        529 GKKQKGKRGGGDEESGEHTQVPADSPDSQEEQKGESSASSPEEPEEITC-----LEKGLA 583

Qy        422 TMRFLSSGKEQKTKPKEKVK------TKPEKPSLPKCSVQVGIKISSVHKRLASEKRENP 475

Db        584 EVQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRV----------RRPSESDKEDE 633

Qy        476 VKKVMLASRSETSGKEAACESSTPSWASDHNYNAVKPEKPEKPTALSPTLLS 527

Db        634 LDKVKSATLSST--------ESTASEMQEEMKGSVEEPKPEEPKRKVDTSVS 677


RESULT 5
US-09-257-179-80
; Sequence 80, Application US/09257179
; Patent No. 6410709
; GENERAL INFORMATION:
;  APPLICANT: Ruben et al.
;  TITLE OF INVENTION: 29 Human Secreted Proteins
;  FILE REFERENCE: PZ015P1
;  CURRENT APPLICATION NUMBER: US/09/257,179
;  CURRENT FILING DATE: 1999-02-25
;  EARLIER APPLICATION NUMBER: PCT/US98/17709
;  EARLIER FILING DATE: 1998-08-27
;  EARLIER APPLICATION NUMBER: 60/056,270
;  EARLIER FILING DATE: 1997-08-29
;  EARLIER APPLICATION NUMBER: 60/056,271
;  EARLIER FILING DATE: 1997-08-29
;  EARLIER APPLICATION NUMBER: 60/056,247
;  EARLIER FILING DATE: 1997-08-29
;  EARLIER APPLICATION NUMBER: 60/056,073
;  EARLIER FILING DATE: 1997-08-29
;  NUMBER OF SEQ ID NOS: 128
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 80
;   LENGTH: 238
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-257-179-80

  Query Match             5.1%;  Score 164;  DB 4;  Length 238;
  Best Local Similarity   36.2%;  Pred. No. 3.4e-06;
  Matches   29;  Conservative   16;  Mismatches   11;  Indels   24;  Gaps    3;

Qy        238 KATQGNTEENPREAGKPKPECEVYDPNALYCICRQPHN-NRFMICCDRCEEWFHGDCVGI 296

Db         51 RAQEGSSEE------------------LYCICRTPYDESQPYIGCDRCQNWYHGRCVGI 91

Qy        297 SEARGRLLERNGEDYICPNC 316

Db         92 LQSEAELI----DEYVCPQC 107


RESULT 6
US-08-296-791-6
; Sequence 6, Application US/08296791
; Patent No. 6245337
;  GENERAL INFORMATION:
;    APPLICANT:  St. Geme III, Joseph W.
;    APPLICANT:  Falkow, Stanley
;    TITLE OF INVENTION:  Haemophilus Adherence and Penetration
;    TITLE OF INVENTION:  Protein
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/296,791
;      FILING DATE:  25-AUG-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  A-59941/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1848 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
US-08-296-791-6

  Query Match             4.8%;  Score 153.5;  DB 3;  Length 1848;
  Best Local Similarity   21.7%;  Pred. No. 0.00055;
  Matches  124;  Conservative   71;  Mismatches  208;  Indels  169;  Gaps   25;

Qy          8 SNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPSE------QQPQQHNLSLRR 61

Db       1072 STETAPKSDTATQTE----------------NPNSESVPSETTEKVAENPPQENETVAK 1114

Qy         62 SGRQ-----PKRTERVEEPLTTV----------RRRGKKNVPVSLEDSSEPTSSTVTDVE 106

Db       1115 NEQEATEPTPQNGEVAKEDQPTVEANTQTNEATQSEGKTEETQTAETKSEPTES-VTVSE 1173

Qy        107 TASEGSVESSSEIRSGPVSDSLGKEHPASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLR 166

Db       1174 NQPEKTVSQSTE-------DKVVVE---KEEKAK--VETEETQKAPQ--VTSKEPPKQAE 1219

Qy        167 RKREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSAEQ-DRPLCKQEPE-------- 217

Db       1220 PAPEEVPTDTNAEEAQALQQTQPTTVAAAETTSPNSKPAEETQQPSEKTNAEPVTPVVSE 1279

Qy        218 --ASQGPVSQSETDDIENQLEGKATQGNTEENPRE---AGKP----KPECEVYDPNALYC 268

Db       1280 NTATQ-PTETEETAKVEKEKTQEVPQVASQESPKQEQPAAKPQAQTKPQAEPARENVL-- 1336

Qy        269 ICRQPHNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGS 328

Db       1337 --------------------------------TTKNVGEPQ-------PQAQPQTQST 1355

Qy        329 A---TNEQDSGCRSVGADGTDC--------TSIGTVEQKSGEDQGIKGRIEKAANPSGKK 377

Db       1356 AVPTTGETAANSKPAAKPQAQAKPQTEPARENVSTVNTKEPQSQ-------TSATVSTEQ 1408

Qy        378 KLKIFQPVVEAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAAATMRFLSSGKEQKTKPK 437

Db       1409 PAKETSSNVEQPAPENSINTG-------------------SATTM---TETAEKSDKPQ 1445

Qy        438 EKVKT----KPEKPSLPKCSVQVGIKIS---SVHKRLASEKRENPVKKVMLASRSETSGK 490

Db       1446 METVTENDRQPEANTVADNSVANNSESSESKSRRRRSVSQPKETSAEETTVASTQETTVD 1505

Qy        491 EAACESSTPSWASDHNYNAVKP---EKPEKPT 519

Db       1506 NSV---STPKPRSRRTRRSVQTNSYEPVELPT 1534


RESULT 7
US-09-839-996-6
; Sequence 6, Application US/09839996
; Patent No. 6642371
;  GENERAL INFORMATION:
;       APPLICANT: St. Geme III, Joseph W.
;                  Falkow, Stanley
```

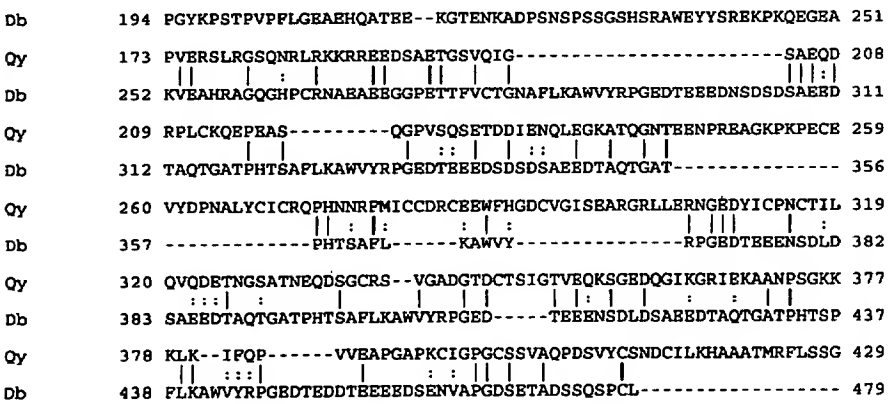
```
;       TITLE OF INVENTION: Haemophilus Adherence and Penetration
;                           Protein
;       NUMBER OF SEQUENCES: 9
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;            STREET: 4 Embarcadero Center, Suite 3400
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States
;            ZIP: 94111-4187
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/839,996
;            FILING DATE: 20-Apr-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/296,791
;            FILING DATE: 25-AUG-1994
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Trecartin, Richard F.
;            REGISTRATION NUMBER: 31,801
;            REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 781-1989
;            TELEFAX: (415) 398-3249
;            TELEX: 910 277299
;   INFORMATION FOR SEQ ID NO: 6:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1848 amino acids
;            TYPE: amino acid
;            TOPOLOGY: unknown
;       SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-839-996-6

  Query Match             4.8%;  Score 153.5;  DB 4;  Length 1848;
  Best Local Similarity   21.7%;  Pred. No. 0.00055;
  Matches  124;  Conservative   71;  Mismatches  208;  Indels  169;  Gaps   25;

Qy          8 SNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPSE------QQPQQHNLSLRR 61
              | | |||:   | |                       ::|: |||      : | | | :: :
Db       1072 STETAPKSDTATQTE----------------NPNSESVPSETTEKVAENPPQENETVAK 1114

Qy         62 SGRQ-----PKRTERVEEFLTTV----------RRRGKKNVPVSLEDSSEPTSSTVTDVE 106
              : ::     |: | :|   ||          :  |  :| |||| | ||  |
Db       1115 NEQEATEPTPQNGEVAKEDQPTVEANTQTNEATQSEGKTEETQTAETKSEPTES-VTVSE 1173

Qy        107 TASEGSVESSSEIRSGPVSDSLGKEHPASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLR 166
              | :|  |:|       | :  |     ||||   | |:|  :    :| ||   :
Db       1174 NQPEKTVSQSTE-------DKVVVE---KEEKAK--VETEETQKAPQ--VTSKEPPKQAE 1219

Qy        167 RKREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSAEQ-DRPLCKQEPE-------- 217
              |: | : :   :|   : :     :||| |      ||:  :|  |   |
Db       1220 PAPEEVPTDTNAEEAQALQQTQPTTVAAAETTSPNSKPAEETQQPSEKTNAEPVTPVVSE 1279

Qy        218 --ASQGPVSQSETDDIENQLEGKATQGNTEENPRE---AGKP----KPECEVYDPNALYC 268
              |:| |    ||  :| :  :  |  ::|:|::   | ||     ||:  |    | |
Db       1280 NTATQ-PTETEETAKVEKEKTQEVPQVASQESPKQEQPAAKPQAQTKPQAEPARENVL-- 1336

Qy        269 ICRQPHNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGS 328
                                                :  ||          | | :|  :
Db       1337 ---------------------------------TTKNVGEPQ-------PQAQPQTQST 1355

Qy        329 A---TNEQDSGCRSVGADGTDC--------TSIGTVEQKSGEDQGIKGRIEKAANPSGKK 377
              |   | |  :  :                 :: ||  |  : |        :|  | ::
Db       1356 AVPTTGETAANSKPAAKPQAQAKPQTEPARENVSTVNTKEPQSQ-------TSATVSTEQ 1408

Qy        378 KLKIFQPVVEAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAAATMRFLSSGKEQKTKPK 437
              |     || |    |  |                :| ||   :   |:  ||:
Db       1409 PAKETSSNVEQPAPENSINTG-------------------SATTM---TETAEKSDKPQ 1445

Qy        438 EKVKT----KPEKFSLPKCSVQVGIKIS---SVHKRLASEKRENPVKKVMLASRSETSGK 490
              :  |    :|| ::   ||    : |   |  :|  |: :|   ::  :|| ||:
Db       1446 METVTENDRQPEANTVADNSVANNSESSESKSRRRRSVSQPKETSAEETTVASTQETTVD 1505

Qy        491 EAACESSTPSWASDHNYNAVKP---EKPEKPT 519
              :  |||   |    :|:   | | ||
Db       1506 NSV---STPKPRSRRTRRSVQTNSYEPVELPT 1534


RESULT 8
US-10-080-505-6
; Sequence 6, Application US/10080505
; Patent No. 6676948
; GENERAL INFORMATION:
;  APPLICANT: St. Geme, Joseph W.
;  TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTIENS
;  FILE REFERENCE: A-59941-1/RFT/DCF/DHR
;  CURRENT APPLICATION NUMBER: US/10/080,505
;  CURRENT FILING DATE:  2002-02-22
;  PRIOR APPLICATION NUMBER: US 08/296,791
;  PRIOR FILING DATE: 1994-10-25
;  PRIOR APPLICATION NUMBER: US 09/839,996
;  PRIOR FILING DATE: 2001-04-20
;  NUMBER OF SEQ ID NOS: 58
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
;   LENGTH: 1848
;   TYPE: PRT
;   ORGANISM: Haemophilus influenzae
US-10-080-505-6

  Query Match             4.8%;  Score 153.5;  DB 4;  Length 1848;
  Best Local Similarity   21.7%;  Pred. No. 0.00055;
  Matches  124;  Conservative   71;  Mismatches  208;  Indels  169;  Gaps   25;

Qy          8 SNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPSE------QQPQQHNLSLRR 61
              | | |||:   | |                       ::|: |||      : | | | :: :
Db       1072 STETAPKSDTATQTE----------------NPNSESVPSETTEKVAENPPQENETVAK 1114

Qy         62 SGRQ-----PKRTERVEEFLTTV----------RRRGKKNVPVSLEDSSEPTSSTVTDVE 106
              : ::     |: | :|   ||          :  |  :| |||| | ||  |
Db       1115 NEQEATEPTPQNGEVAKEDQPTVEANTQTNEATQSEGKTEETQTAETKSEPTES-VTVSE 1173

Qy        107 TASEGSVESSSEIRSGPVSDSLGKEHPASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLR 166
              | :|  |:|       | :  |     ||||   | |:|  :    :| ||   :
Db       1174 NQPEKTVSQSTE-------DKVVVE---KEEKAK--VETEETQKAPQ--VTSKEPPKQAE 1219

Qy        167 RKREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSAEQ-DRPLCKQEPE-------- 217
              |: | : :   :|   : :     :||| |      ||:  :|  |   |
Db       1220 PAPEEVPTDTNAEEAQALQQTQPTTVAAAETTSPNSKPAEETQQPSEKTNAEPVTPVVSE 1279

Qy        218 --ASQGPVSQSETDDIENQLEGKATQGNTEENPRE---AGKP----KPECEVYDPNALYC 268
              |:| |    ||  :| :  :  |  ::|:|::   | ||     ||:  |    | |
Db       1280 NTATQ-PTETEETAKVEKEKTQEVPQVASQESPKQEQPAAKPQAQTKPQAEPARENVL-- 1336

Qy        269 ICRQPHNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGS 328
                                                :  ||          | | :|  :
Db       1337 ---------------------------------TTKNVGEPQ-------PQAQPQTQST 1355

Qy        329 A---TNEQDSGCRSVGADGTDC--------TSIGTVEQKSGEDQGIKGRIEKAANPSGKK 377
              |   | |  :  :                 :: ||  |  : |        :|  | ::
Db       1356 AVPTTGETAANSKPAAKPQAQAKPQTEPARENVSTVNTKEPQSQ-------TSATVSTEQ 1408

Qy        378 KLKIFQPVVEAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAAATMRFLSSGKEQKTKPK 437
              |     || |    |  |                :| ||   :   |:  ||:
Db       1409 PAKETSSNVEQPAPENSINTG-------------------SATTM---TETAEKSDKPQ 1445

Qy        438 EKVKT----KPEKFSLPKCSVQVGIKIS---SVHKRLASEKRENPVKKVMLASRSETSGK 490
              :  |    :|| ::   ||    : |   |  :|  |: :|   ::  :|| ||:
Db       1446 METVTENDRQPEANTVADNSVANNSESSESKSRRRRSVSQPKETSAEETTVASTQETTVD 1505

Qy        491 EAACESSTPSWASDHNYNAVKP---EKPEKPT 519
```

```
                :    |||   |     :|:    |  | ||
Db       1506 NSV---STPKPRSRRTRRSVQTNSYEPVELPT 1534


RESULT 9
PCT-US95-10661A-6
; Sequence 6, Application PC/TUS9510661A
;  GENERAL INFORMATION:
;    APPLICANT:  Washington University, et al.
;    TITLE OF INVENTION:  Haemophilus Adherence and Penetration Protein
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/10661A
;      FILING DATE:  16-AUG-1995
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/296,791
;      FILING DATE:  25-AUG-1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  FP-59941/RFT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1848 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
PCT-US95-10661A-6

  Query Match             4.8%;  Score 153.5;  DB 5;  Length 1848;
  Best Local Similarity   21.7%;  Pred. No. 0.00055;
  Matches  124;  Conservative   71;  Mismatches  208;  Indels  169;  Gaps   25;

Qy          8 SNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPSE------QQPQQHNLSLRR 61
              | | |||:  |  |                    ::|: |||      : | | | :: :
Db       1072 STETAPKSDTATQTE----------------NPNSESVPSETTEKVAENPPQENETVAK 1114

Qy         62 SGRQ-----PKRTERVEEFLTTV----------RRRGKKNVPVSLEDSSEPTSSTVTDVE 106
              : ::     |:  |  :|   ||          :  ||     : |  |||| | ||  |
Db       1115 NEQEATEPTPQNGEVAKEDQPTVEANTQTNEATQSEGKTEETQTAETKSEPTES-VTVSE 1173

Qy        107 TASEGSVESSSEIRSGPVSDSLGKEHPASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLR 166
              | :|  |:|       | :  |     ||||   | |:|  :    :| ||   :
Db       1174 NQPEKTVSQSTE-------DKVVVE---KEEKAK--VETEETQKAPQ--VTSKEPPKQAE 1219

Qy        167 RKREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSAEQ-DRPLCKQEPE-------- 217
              |: | : :   :|   : :     :||| |      ||:  :|  |   |
Db       1220 PAPEEVPTDTNAEEAQALQQTQPTTVAAAETTSPNSKPAEETQQPSEKTNAEPVTPVVSE 1279

Qy        218 --ASQGPVSQSETDDIENQLEGKATQGNTEENPRE---AGKP----KPECEVYDPNALYC 268
                |:| |    ||  :| :   :  |  ::|:|::   | ||    ||: |     | |
Db       1280 NTATQ-PTETEETAKVEKEKTQEVPQVASQESPKQEQPAAKPQAQTKPQAEPARENVL-- 1336

Qy        269 ICRQPHNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGS 328
                                          :  ||          | | :|  :
Db       1337 --------------------------------TTKNVGEPQ-------PQAQPQTQST 1355
```

```
Qy        329 A---TNEQDSGCRSVGADGTDC--------TSIGTVEQKSGEDQGIKGRIEKAANPSGKK 377
              |   | ||  :  :                    :: ||  |  : |         :|  | ::
Db       1356 AVPTTGETAANSKPAAKPQAQAKPQTEPARENVSTVNTKEPQSQ-------TSATVSTEQ 1408

Qy        378 KLKIFQPVVEAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAAATMRFLSSGKEQKTKPK 437
                |     || |      |  |                      :| ||   :   |:  ||:
Db       1409 PAKETSSNVEQPAPENSINTG-------------------SATTM---TETAEKSDKPQ 1445

Qy        438 EKVKT----KPEKPSLPKCSVQVGIKIS---SVHKRLASEKRENPVKKVMLASRSETSGK 490
              :  |    :|| ::   ||    : |   |  :|  |: :|   ::  :||  ||:
Db       1446 METVTENDRQPEANTVADNSVANNSESSESKSRRRRSVSQPKETSAEETTVASTQETTVD 1505

Qy        491 EAACESSTPSWASDHNYNAVKP---EKPEKPT 519
              :   |||   |     :|:    |  | ||
Db       1506 NSV---STPKPRSRRTRRSVQTNSYEPVELPT 1534


RESULT 10
US-08-285-440-5
; Sequence 5, Application US/08285440
; Patent No. 5532337
;  GENERAL INFORMATION:
;    APPLICANT:  Ken'ichiro HAYASHI et al.
;    TITLE OF INVENTION:  POLYPEPTIDE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack
;      STREET:  805 Fifteenth Street, N.W., #700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 500 kb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  DisplayWrite
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/285,440
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/858,947
;      FILING DATE:  March 27, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-8850
;      TELEFAX:
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  532 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    HYPOTHETICAL:
;    ANTI-SENSE:
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
;      ORGANISM:
;      STRAIN:
;      INDIVIDUAL ISOLATE:
;      DEVELOPMENTAL STAGE:
;      HAPLOTYPE:
;      TISSUE TYPE:
;      CELL TYPE:
;      CELL LINE:
;      ORGANELLE:
```

```
;     IMMEDIATE SOURCE:
;       LIBRARY:
;       CLONE:
;     POSITION IN GENOME:
;       CHROMOSOME/SEGMENT:
;       MAP POSITION:
;       UNITS:
;     FEATURE:
;       NAME/KEY:
;       LOCATION:
;       IDENTIFICATION METHOD:
;       OTHER INFORMATION:
;     PUBLICATION INFORMATION:
;       AUTHORS:
;       TITLE:
;       JOURNAL:
;       VOLUME:
;       ISSUE:
;       PAGES:
;       DATE:
;       DOCUMENT NUMBER:
;       FILING DATE:
;       PUBLICATION DATE:
;       RELEVANT RESIDUES IN SEQ ID NO:
US-08-285-440-5

  Query Match             4.7%;  Score 151.5;  DB 1;  Length 532;
  Best Local Similarity   20.9%;  Pred. No. 0.00013;
  Matches  120;  Conservative   66;  Mismatches  233;  Indels  155;  Gaps   23;

Qy         30 RRTTIAKREGAGDTEADPSEQQPQQHNLSLRRSGRQPK-RTERVEEFLTTVRRRGKKNVP 88
              ||:  |  : |      |  |:  ::     ||  || : | :: || |  |  : : |
Db         10 RRSLAALSQIAYQRNDDDEEEAARER----RRRARQERLRQKQEEESLGQVTDQVEVNAQ 65

Qy         89 VSLEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDSLGKEHPASSEKAKGGEEEEDT 148
               |: |    |::| | ||                    | :  |  |:    ||
Db         66 NSVPDEEAKTTTTNTQVE--------------------GDDEAAFLERLARREERRQ- 102

Qy        149 SDSDSDGLTLKELQNRLRRKREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSAEQD 208
                        | ||  | |::| :|   ::  :   |  :| : |:||      : |::
Db        103 ----------KRLQEALERQKEFDP---TITDASLSLPSRRMQNDTAEN-----ETTEKE 144

Qy        209 RPLCKQEPEASQGPVSQSET-----------DDIENQLEGKATQGNTEENPREAG----- 252
                | |    :  : ::||            |  ||: | |  :   || |:
Db        145 E---KSESRQERYEIEETETVTKSYQKNDWRDAEENKKEDKEKEEEEEEKPKRGSIGENQ 201

Qy        253 ----------KPKPE-----------CEVYDPNALYCICRQPH-NNRFMICCDRCEEWFH 290
                       :|| |            ||   |  :   :  | | |      |
Db        202 IKDEKIKKDKEPKEEVKSFMDRKKGFTEVKSQNGEFMTHKLKHTENTFSRPGGRAS---- 257

Qy        291 GDCVGISEARG-----------RLLERNGE------------DYICPNCTILQVQDETN 326
                |   || |            |  | ||             |: : |
Db        258 ---VDTKRAEGAPQVEAGKRLEELRRRRGETESEEPEKLKQKQQEAALELEELKKKREER 314

Qy        327 GSATNEQDSGCRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLKIFQPVV 386
                |::   :   ||          ::  | : :|  ||:    : :|: |: :  :
Db        315 RKVLEEEEQRRKQEEAD--------RKLREEEEKRRLKEEIERRRAEAAEKRQKMPEDGL 366

Qy        387 EAPGAP-KCIGPGCSSVAQPDSVYCSNDCILK-------HAAATMRFLSSGKEQKTKPKE 438
                | ||  |  ||:   :    |  : |       | ||  :  : |  ||  |   |
Db        367 SDDKKPFKCFTPKGSSLKIEERAEFLNKSVQKSSGVKSTHQAAIVSKIDSRLEQYTSAIE 426

Qy        439 KVK----TKPEKPSLPKCSVQVGIKISSVHKRLASEKRENPVKKVMLAS--RSETSGKEA 492
               |    |||    ||       :    |    :  ::  |        |     ||:| :
Db        427 GTKSAKPTKPAASDLP-------VPAEGVRNIKSMWEKGNVFSSPTAAGTPNKETAGLKV 479

Qy        493 ACESSTPSWAS---DHNYNAVKPEKPEKPTALSP 523
                |    | :   | |    |   | ||: | |
Db        480 GVSSRINEWLTKTPDGN----KSPAP-KPSDLRP 508


RESULT 11
US-08-630-349-5
; Sequence 5, Application US/08630349
; Patent No. 5739008
; GENERAL INFORMATION:
;   APPLICANT:  Ken'ichiro HAYASHI et al.
;   TITLE OF INVENTION:  POLYPEPTIDE
;   NUMBER OF SEQUENCES:  32
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Wenderoth, Lind & Ponack
;     STREET:  805 Fifteenth Street, N.W., #700
;     CITY:  Washington
;     STATE:  D.C.
;     COUNTRY:  U.S.A.
;     ZIP:  20005
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Diskette, 3.5 inch, 144 mb
;     COMPUTER:  IBM Compatible
;     OPERATING SYSTEM:  MS-DOS
;     SOFTWARE:  Wordperfect 5.1
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/630,349
;     FILING DATE:  April 10, 1996
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  08/285,440
;     FILING DATE:  August 4, 1994
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  07/858,947
;     FILING DATE:  March 27, 1992
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Warren M. Cheek, Jr.
;     REGISTRATION NUMBER:  33,367
;  INFORMATION FOR SEQ ID NO:  5:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  532 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  peptide
US-08-630-349-5

  Query Match             4.7%;  Score 151.5;  DB 1;  Length 532;
  Best Local Similarity   20.9%;  Pred. No. 0.00013;
  Matches  120;  Conservative   66;  Mismatches  233;  Indels  155;  Gaps   23;

Qy         30 RRTTIAKREGAGDTEADPSEQQPQQHNLSLRRSGRQPK-RTERVEEFLTTVRRRGKKNVP 88
              ||:  |  : |      |  |:  ::     ||  || : | :: || |  |  : : |
Db         10 RRSLAALSQIAYQRNDDDEEEAARER----RRRARQERLRQKQEEESLGQVTDQVEVNAQ 65

Qy         89 VSLEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDSLGKEHPASSEKAKGGEEEEDT 148
               |: |    |::| | ||                    | :  |  |:    ||
Db         66 NSVPDEEAKTTTTNTQVE--------------------GDDEAAFLERLARREERRQ- 102

Qy        149 SDSDSDGLTLKELQNRLRRKREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSAEQD 208
                        | ||  | |::| :|   ::  :   |  :| : |:||      : |::
Db        103 ----------KRLQEALERQKEFDP---TITDASLSLPSRRMQNDTAEN-----ETTEKE 144

Qy        209 RPLCKQEPEASQGPVSQSET-----------DDIENQLEGKATQGNTEENPREAG----- 252
                | |    :  : ::||            |  ||: | |  :   || |:
Db        145 E---KSESRQERYEIEETETVTKSYQKNDWRDAEENKKEDKEKEEEEEEKPKRGSIGENQ 201

Qy        253 ----------KPKPE-----------CEVYDPNALYCICRQPH-NNRFMICCDRCEEWFH 290
                       :|| |            ||   |  :   :  | | |      |
Db        202 IKDEKIKKDKEPKEEVKSFMDRKKGFTEVKSQNGEFMTHKLKHTENTFSRPGGRAS---- 257

Qy        291 GDCVGISEARG-----------RLLERNGE------------DYICPNCTILQVQDETN 326
                |   || |            |  | ||             |: : |
Db        258 ---VDTKRAEGAPQVEAGKRLEELRRRRGETESEEPEKLKQKQQEAALELEELKKKREER 314

Qy        327 GSATNEQDSGCRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLKIFQPVV 386
                |::   :   ||          ::  | : :|  ||:    : :|: |: :  :
Db        315 RKVLEEEEQRRKQEEAD--------RKLREEEEKRRLKEEIERRRAEAAEKRQKMPEDGL 366

Qy        387 EAPGAP-KCIGPGCSSVAQPDSVYCSNDCILK-------HAAATMRFLSSGKEQKTKPKE 438
```

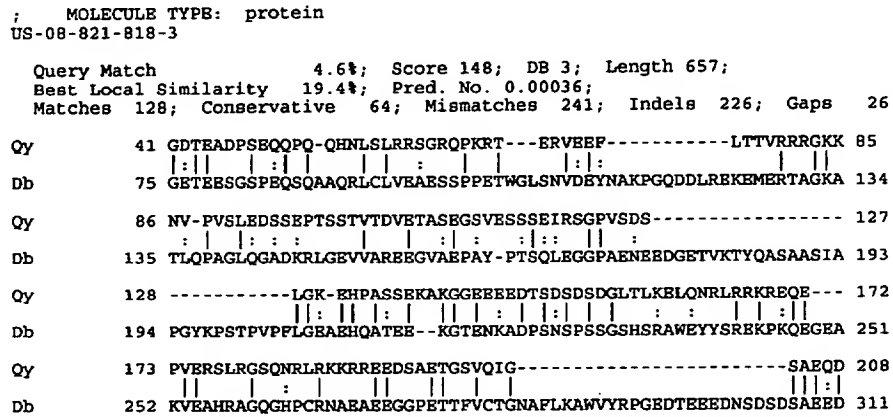
```
                 | ||  |  ||:   :     |  : |       | || :  : |  || |   |
Db        367 SDDKKPFKCFTPKGSSLKIEERAEFLNKSVQKSSGVKSTHQAAIVSKIDSRLEQYTSAIE 426

Qy        439 KVK----TKPEKFSLPKCSVQVGIKISSVHKRLASEKRENPVKKVMLAS--RSETSGKEA 492
                |    |||    ||       :    |    :  :: |       |     ||:| :
Db        427 GTKSAKPTKPAASDLP-------VPAEGVRNIKSMWEKGNVFSSPTAAGTPNKETAGLKV 479

Qy        493 ACESSTPSWAS---DHNYNAVKPEKPEKPTALSP 523
                  |   | :   | |   |  | | ||: | |
Db        480 GVSSRINEWLTKTPDGN----KSPAP-KPSDLRP 508


RESULT 12
US-09-976-594-726
; Sequence 726, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
;  APPLICANT: Furness, Michael
;  APPLICANT:  Buchbinder, Jenny
;  TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
;  FILE REFERENCE: PA-0041 US
;  CURRENT APPLICATION NUMBER: US/09/976,594
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: 60/240,409
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 1143
;  SOFTWARE: PERL Program
; SEQ ID NO 726
;   LENGTH: 2468
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6673549 4295277CD1
US-09-976-594-726

  Query Match             4.6%;  Score 149.5;  DB 4;  Length 2468;
  Best Local Similarity   20.5%;  Pred. No. 0.0018;
  Matches  134;  Conservative   87;  Mismatches  250;  Indels  183;  Gaps   31;

Qy          7 LSNEEAPKAIKPT-----SKEFRKTWGFRRTT---IAKREGAGDTEADPSEQQPQQHNLS 58
              : :||  |  :|        ||   | |   :     |   || |: |  | | :|
Db        861 IEDEEKLKETEPVEAYVIQKEREVTKGPAESPDEGITTTEGEGECEQTPEELEP------ 914

Qy         59 LRRSGRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSE 118
              : : |       : :|:|       .  ::      |:|||                |  |  :|
Db        915 VEKQG-----VDDIEKF---------EDEGAGFEESSE-------------TGDYEEKAE 947

Qy        119 IRSGPVSDSLGKEHPASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLRRKRE-------- 170
                  :  |:||   |        |:|::: :::|           :| |||
Db        948 TEEAEEPEEDGEEHVCVSASKHSPTEDEESAKAEADAY--------IREKRESVASGDDR 999

Qy        171 -QEPVERSL-RGSQNRLRKKRREEDSAETG----------------------SVQIGSAE 206
               :| :: :: :|   :  ::  ||| ||                        : : | ||
Db       1000 AEEDMDEAIEKGEAEQSEEEADEEDKAEDAREEEYEPEKMEAEDYVMAVVDKAAEAGGAE 1059

Qy        207 QD--------RPLCKQ----EPEAS-------QGPVSQSETDDIENQ----LEGKATQGN 243
              :          : |  |     || :|       |  |::   | ||:     |  || |
Db       1060 EQYGFLTTPTKQLGAQSPGREPASSIHDETLPGGSESEATASDEENREDQPEEFTATSGY 1119

Qy        244 TEENPREAGKPKPECEVYDPNALYCICRQPHNN-------RFMICCDRCEEWFHGD---- 292
               |:     : :| |  |:  |      :      ||        :   :   : |   :
Db       1120 TQSTIEISSEPTPMDEMSTPRD---VMSDETNNEETESPSQEFVNITKYESSLYSQEYSK 1176

Qy        293 ------CVGISEARGRLLERNGEDYICPNCTIL--QVQDETNGSATNEQDSGCRSVGADG 344
                   | ||   :    :|:||     ||      :|   | :  :|: | | | |
Db       1177 PADVTPLNGFSEG-SKTDATDGKDYNASASTISPPSSMEEDKFSRSALRDAYCSEVKAST 1235

Qy        345 T----DCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLKIFQPVVEAPGAPKCIGPGCS 400
              |    |  |  : |: |      |      : ||  :|  : :  |      |  |
Db       1236 TLDIKDSISAVSSEKVSPS----KSPSLSPSPPSPLEKTPLGERSVNFSLTPNEIKVSAE 1291

Qy        401 SVAQPDSVYCSNDCILKHAAA----TMRFLS---------------SGKEQKTK--PKE 438
                 :  | |   : : : :| |:    |:  :|                    |  ::|:   | |
Db       1292 AEVAPVSPEVTQEVVEEHCASPEDKTLEVVSPSQSVTGSAGHTPYYQSPTDEKSSHLPTE 1351

Qy        439 KVKTKPEKFSLPKCSVQVGIKISSV--HKRLAS--------EKRENPVKKVMLASRS-ET 487
                ::  |          :| |  : |          ||            |:|::||:   ||
Db       1352 VIEKPP--------AVPVSFEFSDAKDENERASVSPMDEPVPDSESPIEKVLSPLRSPPL 1403

Qy        488 SGKEAACES--STPSWASDHNYNAVKPEKPEK---PTALSPT--LLSKCTYHPK 534
                | |:| ||  |    ||      :   ||  |   | :||   : |   |  |
Db       1404 IGSESAYESFLSADDKASGRGAESPFEEKSGKQGSPDQVSPVSEMTSTSLYQDK 1457


RESULT 13
US-08-893-852A-3
; Sequence 3, Application US/08893852A
; Patent No. 6080558
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Shah, Purvi
;    APPLICANT:  Corley, Neil C.
;    TITLE OF INVENTION:  NEW HUMAN GROWTH REGULATOR PROTEIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/893,852A
;      FILING DATE:  Filed Herewith
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0341 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  657 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  53041
US-08-893-852A-3

  Query Match             4.6%;  Score 148;  DB 3;  Length 657;
  Best Local Similarity   19.4%;  Pred. No. 0.00036;
  Matches  128;  Conservative   64;  Mismatches  241;  Indels  226;  Gaps   26;

Qy         41 GDTEADPSEQQPQ-QHNLSLRRSGRQPKRT---ERVEEF-----------LTTVRRRGKK 85
              |:||   | :| |    | |  :   |  |     |:|:                |  ||
Db         75 GETEESGSPEQSQAAQRLCLVEAESSPPETWGLSNVDEYNAKPGQDDLREKEMERTAGKA 134

Qy         86 NV-PVSLEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDS----------------- 127
                : |  |: : :      |   |  :| :   :|::  ||  :
Db        135 TLQPAGLQGADKRLGEVVAREEGVAEPAY-PTSQLEGGPAENEEDGETVKTYQASAASIA 193

Qy        128 -----------LGK-EHPASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLRRKREQE--- 172
                         ||: || |: |  || | : | |:| | |   :  :   | | :||
```



```
Db        194 PGYKPSTPVPPLGEAEHQATEE--KGTENKADPSNSPSSGSHSRAWEYYSREKPKQEGEA 251

Qy        173 PVERSLRGSQNRLRKKRREEDSAETGSVQIG----------------------SAEQD 208

Db        252 KVEAHRAGQGHPCRNAEAEEGGPETTFVCTGNAFLKAWVYRPGEDTEEEDNSDSDSAEED 311

Qy        209 RPLCKQEPEAS---------QGPVSQSETDDIENQLEGKATQGNTEENPREAGKPKPECE 259

Db        312 TAQTGATPHTSAFLKAWVYRPGEDTEEEDSDSDSAEEDTAQTGAT--------------- 356

Qy        260 VYDPNALYCICRQPHNNRPMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTIL 319

Db        357 -------------PHTSAPL------KAWVY--------------RPGEDTEEENSDLD 382

Qy        320 QVQDETNGSATNEQDSGCRS--VGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKK 377

Db        383 SAEEDTAQTGATPHTSAFLKAWVYRPGED-----TEEENSDLDSAEEDTAQTGATPHTSP 437

Qy        378 KLK--IFQP------VVEAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAAATMRFLSSG 429

Db        438 FLKAWVYRPGEDTEDDTEEEEDSENVAPGDSETADSSQSPCL---------------- 479

Qy        430 KEQKTKPKEKVK----------------TKPEK-FSLPKCSV---------------- 454

Db        480 QPQRCLPGEKTKGRGEEPPLFQVAFYLPGEKPESPWAAPKLPLRLQRRLRLFKAPTRDQD 539

Qy        455 -QVGIKISSVH-------------KRLASEKRENPVKKV---------MLASRSETSGKE 491

Db        540 PEIPLKARKVHFAEKVTVHFLAVWAGPAQAARRGPWEQFARDRSRFARRIAQAEEKLGPY 599

Qy        492 AACESSTPSWASDHNYNAVKPEKPEKPTALSPTLLSKCTYHP-KAGFPGPSHHLGGCLG 549

Db        600 LTPDSRARAWARLRNPSLPQSE-PRSSSEATPLTQDVTTPSPLPSETPSPSLYLGGRRG 657
```

RESULT 14
US-08-821-818-3
; Sequence 3, Application US/08821818
; Patent No. 6146877
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; TITLE OF INVENTION: IDENTIFICATION OF THE PROGRESSION
; TITLE OF INVENTION: ELEVATED GENE-3 AND USES THEREOF
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/821,818
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Chan, Albert Wai-Kit
; REGISTRATION NUMBER: 36,479
; REFERENCE/DOCKET NUMBER: 51523
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 657 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-821-818-3

Query Match 4.6%; Score 148; DB 3; Length 657;
Best Local Similarity 19.4%; Pred. No. 0.00036;
Matches 128; Conservative 64; Mismatches 241; Indels 226; Gaps 26;



```
Qy         41 GDTEADPSEQQPQ-QHNLSLRRSGRQPKRT---ERVEEF----------LTTVRRRGKK 85

Db         75 GETEESGSPEQSQAAQRLCLVEAESSPPETWGLSNVDEYNAKPGQDDLREKEMERTAGKA 134

Qy         86 NV-PVSLEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDS---------------- 127

Db        135 TLQPAGLQGADKRLGEVVAREEGVAEPAY-PTSQLEGGPAENEEDGETVKTYQASAASIA 193

Qy        128 ----------LGK-EHPASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLRRKREQE--- 172

Db        194 PGYKPSTPVPPLGEAEHQATEE--KGTENKADPSNSPSSGSHSRAWEYYSREKPKQEGEA 251

Qy        173 PVERSLRGSQNRLRKKRREEDSAETGSVQIG----------------------SAEQD 208

Db        252 KVEAHRAGQGHPCRNAEAEEGGPETTFVCTGNAFLKAWVYRPGEDTEEEDNSDSDSAEED 311

Qy        209 RPLCKQEPEAS---------QGPVSQSETDDIENQLEGKATQGNTEENPREAGKPKPECE 259

Db        312 TAQTGATPHTSAFLKAWVYRPGEDTEEEDSDSDSAEEDTAQTGAT--------------- 356

Qy        260 VYDPNALYCICRQPHNNRPMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTIL 319

Db        357 -------------PHTSAPL------KAWVY--------------RPGEDTEEENSDLD 382

Qy        320 QVQDETNGSATNEQDSGCRS--VGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKK 377

Db        383 SAEEDTAQTGATPHTSAFLKAWVYRPGED-----TEEENSDLDSAEEDTAQTGATPHTSP 437

Qy        378 KLK--IFQP------VVEAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAAATMRFLSSG 429

Db        438 FLKAWVYRPGEDTEDDTEEEEDSENVAPGDSETADSSQSPCL---------------- 479

Qy        430 KEQKTKPKEKVK----------------TKPEK-FSLPKCSV---------------- 454

Db        480 QPQRCLPGEKTKGRGEEPPLFQVAFYLPGEKPESPWAAPKLPLRLQRRLRLFKAPTRDQD 539

Qy        455 -QVGIKISSVH-------------KRLASEKRENPVKKV---------MLASRSETSGKE 491

Db        540 PEIPLKARKVHFAEKVTVHFLAVWAGPAQAARRGPWEQFARDRSRFARRIAQAEEKLGPY 599

Qy        492 AACESSTPSWASDHNYNAVKPEKPEKPTALSPTLLSKCTYHP-KAGFPGPSHHLGGCLG 549

Db        600 LTPDSRARAWARLRNPSLPQSE-PRSSSEATPLTQDVTTPSPLPSETPSPSLYLGGRRG 657
```

RESULT 15
US-09-052-753B-3
; Sequence 3, Application US/09052753B
; Patent No. 6472520
; GENERAL INFORMATION:
; APPLICANT: Paul B. Fisher
; TITLE OF INVENTION: Progression Elevated Gene-3 and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: A34608-B
; CURRENT APPLICATION NUMBER: US/09/052,753B
; CURRENT FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: PCT/US98/05793
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 08/812,818
; PRIOR FILING DATE: 1997-03-21
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 657
; TYPE: PRT
; ORGANISM: rat

```
US-09-052-753B-3

  Query Match             4.6%;  Score 148;  DB 4;  Length 657;
  Best Local Similarity   19.4%;  Pred. No. 0.00036;
  Matches  128;  Conservative   64;  Mismatches  241;  Indels  226;  Gaps   26;

Qy         41 GDTEADPSEQQPQ-QHNLSLRRSGRQPKRT---ERVEEF-----------LTTVRRRGKK 85

Db         75 GETEESGSPEQSQAAQRLCLVEAESSPPETWGLSNVDEYNAKPGQDDLREKEMERTAGKA 134

Qy         86 NV-PVSLEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDS----------------- 127

Db        135 TLQPAGLQGADKRLGEVVAREEGVAEPAY-PTSQLEGGPAENEEDGETVKTYQASAASIA 193

Qy        128 -----------LGK-EHPASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLRRKREQE--- 172

Db        194 PGYKPSTPVPFLGEAEHQATEE--KGTENKADPSNSPSSGSHSRAWEYYSREKPKQEGEA 251

Qy        173 PVERSLRGSQNRLRKKRREEDSAETGSVQIG-----------------------SAEQD 208

Db        252 KVEAHRAGQGHPCRNAEAEEGGPETTFVCTGNAFLKAWVYRPGEDTEEEDNSDSDSAEED 311

Qy        209 RPLCKQEPEAS---------QGPVSQSETDDIENQLEGKATQGNTEENPREAGKPKPECE 259

Db        312 TAQTGATPHTSAFLKAWVYRPGEDTEEEDSDSDSAEEDTAQTGAT--------------- 356

Qy        260 VYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTIL 319

Db        357 -------------PHTSAFL------KAWVY---------------RPGEDTEEENSDLD 382

Qy        320 QVQDETNGSATNEQDSGCRS--VGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKK 377

Db        383 SAEEDTAQTGATPHTSAFLKAWVYRPGED-----TEEENSDLDSAEEDTAQTGATPHTSP 437

Qy        378 KLK--IFQP------VVEAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAAATMRFLSSG 429

Db        438 FLKAWVYRPGEDTEDDTEEEEDSENVAPGDSETADSSQSPCL------------------ 479

Qy        430 KEQKTKPKEKVK-----------------TKPEK-FSLPKCSV----------------- 454

Db        480 QPQRCLPGEKTKGRGEEPPLFQVAFYLPGEKPESPWAAPKLPLRLQRRLRLFKAPTRDQD 539

Qy        455 -QVGIKISSVH-------------KRLASEKRENPVKKV---------MLASRSETSGKE 491

Db        540 PEIPLKARKVHFAEKVTVHFLAVWAGPAQAARRGPWEQFARDRSRFARRIAQAEEKLGPY 599

Qy        492 AACESSTPSWASDHNYNAVKPEKPEKPTALSPTLLSKCTYHP-KAGFPGPSHHLGGCLG 549

Db        600 LTPDSRARAWARLRNPSLPQSE-PRSSSEATPLTQDVTTPSPLPSETPSPSLYLGGRRG 657

Search completed: April 20, 2004, 21:20:58
Job time : 25 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 20, 2004, 20:57:44 ; Search time 60 Seconds
(without alignments)
2891.403 Million cell updates/sec

Title: US-09-787-016A-4
Perfect score: 3228
Sequence: 1 MDDKGHLSNEEAPKAIKPTS.....LKSCVGLMLEAISYFSFRPW 614

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3228	100.0	614	3	AAY67580	Aay67580 Murine de
2	2110	65.4	562	4	AAB93638	Aab93638 Human pro
3	2108	65.3	562	3	AAY67579	Aay67579 Human dea
4	1792.5	55.5	1191	4	AAM80219	Aam80219 Human pro
5	1756.5	54.4	775	4	ABG22389	Abg22389 Novel hum
6	1693.5	52.5	647	4	AAM79235	Aam79235 Human pro
7	1224.5	37.9	330	4	ABG22388	Abg22388 Novel hum
8	740	22.9	181	3	AAB43724	Aab43724 Human can
9	732	22.7	167	4	ABG22387	Abg22387 Novel hum
10	559	17.3	108	5	ABP05496	Abp05496 Human ORF
11	323.5	10.0	2016	4	ABB71487	Abb71487 Drosophil
12	239.5	7.4	1728	4	AAM39234	Aam39234 Human pol
13	239.5	7.4	1766	4	AAM79892	Aam79892 Human pro
14	239.5	7.4	1766	4	AAM41020	Aam41020 Human pol
15	239.5	7.4	2039	6	ABR64258	Abr64258 Angiogene
16	239.5	7.4	2039	7	ADD45832	Add45832 Human Pro
17	234	7.2	1539	4	AAM78908	Aam78908 Human pro
18	209	6.5	663	4	ABB70523	Abb70523 Drosophil
19	200.5	6.2	3238	4	ABB71715	Abb71715 Drosophil
20	200	6.2	386	4	ABB70522	Abb70522 Drosophil
21	197.5	6.1	1805	2	AAR27204	Aar27204 Rat nesti
22	197.5	6.1	1805	2	AAR60126	Aar60126 Rat nesti
23	197.5	6.1	1805	7	ADD48689	Add48689 Rat Prote
24	194.5	6.0	676	4	AAM47218	Aam47218 Human NOV
25	194.5	6.0	676	5	ABB05039	Abb05039 Human NOV
26	191.5	5.9	676	4	AAG64373	Aag64373 Human hep
27	191.5	5.9	676	4	AAE09441	Aae09441 Human SB-
28	191.5	5.9	717	4	AAE09442	Aae09442 Human SB-
29	190.5	5.9	518	3	AAB53322	Aab53322 Human col
30	190.5	5.9	667	5	ABB05040	Abb05040 Human NOV
31	190.5	5.9	667	5	ABB05041	Abb05041 Human NOV
32	188.5	5.8	555	4	AAG75073	Aag75073 Human col
33	188.5	5.8	671	3	AAY99426	Aay99426 Human PRO
34	188.5	5.8	671	4	AAB66175	Aab66175 Protein o
35	188.5	5.8	671	4	AAU29196	Aau29196 Human PRO
36	188.5	5.8	671	5	ABB05042	Abb05042 Human NOV
37	188.5	5.8	671	6	ABU58572	Abu58572 Human PRO
38	188.5	5.8	671	6	ABU88120	Abu88120 Novel hum
39	188.5	5.8	671	6	ABU84435	Abu84435 Human sec
40	188.5	5.8	671	6	ABR66309	Abr66309 Human sec
41	188.5	5.8	671	6	ABR65699	Abr65699 Human sec
42	188.5	5.8	671	6	ABU99639	Abu99639 Human sec
43	188.5	5.8	671	6	ABU82878	Abu82878 Human PRO
44	188.5	5.8	671	6	ABU89999	Abu89999 Novel hum
45	188.5	5.8	671	6	ABR68248	Abr68248 Human sec

ALIGNMENTS

RESULT 1
AAY67580
ID AAY67580 standard; protein; 614 AA.
XX
AC AAY67580;
XX
DT 19-JUN-2000 (first entry)
XX
DE Murine death inducer-obliterator 1 (DIO-1) polypeptide.
XX
KW Death inducer-obliterator 1; DIO-1; cell death; cancer; tumour; mouse;
KW autoimmune disease; cytostatic; immunosuppressive; antidiabetic;
KW antirheumatic; antiinflammatory; antiproliferative.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT Region 162. .170
FT /note= "NLS sequence"
FT Region 182. .190
FT /note= "NLS sequence"
FT Region 262. .285
FT /note= "zinc finger motif"
FT Region 290. .315
FT /note= "zinc finger motif"
XX
PN WO200015787-A1.
XX
PD 23-MAR-2000.
XX
PF 10-SEP-1999; 99WO-GB003019.
XX
PR 10-SEP-1998; 98SE-00003069.
PR 17-SEP-1998; 98US-0100873P.
XX
PA (CNSJ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.
PA (BANN/) BANNERMAN D G.
XX
PI Alonso CM, Domingo DG, Grandien A, Leonardo E, Martinez P;
XX
DR WPI; 2000-271426/23.
DR N-PSDB; AAZ90579.
XX
PT New DNA encoding human and murine death inducer-obliterator 1
PT polypeptides, useful in the treatment of cancer, autoimmune diseases,
PT diabetes, rheumatoid arthritis, benign tumors, malignant tumors and
PT hyperproliferative skin disorders.
XX

```
PS   Claim 8; Fig 1D; 27pp; English.
XX
CC   The invention provides nucleic acids encoding the human and murine death
CC   inducer-obliterator 1 (DIO-1) polypeptides. The polypeptides can be
CC   expressed by standard recombinant methodology. The DIO-1 polypeptides,
CC   agonists and antagonists are used as a medicament for treating diseases
CC   characterized by an alteration in cell death or by hyperproliferation,
CC   e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign
CC   tumours, malignant tumours or hyperproliferative skin disorders. They are
CC   also useful in the treatment of metabolic, proliferative or inflammatory
CC   conditions. The present sequence represents the murine DIO-1 polypeptide
XX
SQ   Sequence 614 AA;

  Query Match             100.0%;  Score 3228;  DB 3;  Length 614;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-215;
  Matches  614;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MDDKGHLSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPSEQQPQQHNLSLR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MDDKGHLSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPSEQQPQQHNLSLR 60

Qy         61 RSGRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSEIR 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RSGRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSEIR 120

Qy        121 SGPVSDSLGKEHPASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLRRKREQEPVERSLRG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SGPVSDSLGKEHPASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLRRKREQEPVERSLRG 180

Qy        181 SQNRLRKKRREEDSAETGSVQIGSAEQDRPLCKQEPEASQGPVSQSETDDIENQLEGKAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SQNRLRKKRREEDSAETGSVQIGSAEQDRPLCKQEPEASQGPVSQSETDDIENQLEGKAT 240

Qy        241 QGNTEENPREAGKPKPECEVYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEAR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QGNTEENPREAGKPKPECEVYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEAR 300

Qy        301 GRLLERNGEDYICPNCTILQVQDETNGSATNEQDSGCRSVGADGTDCTSIGTVEQKSGED 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GRLLERNGEDYICPNCTILQVQDETNGSATNEQDSGCRSVGADGTDCTSIGTVEQKSGED 360

Qy        361 QGIKGRIEKAANPSGKKKLKIFQPVVEAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 QGIKGRIEKAANPSGKKKLKIFQPVVEAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAA 420

Qy        421 ATMRFLSSGKEQKTKPKEKVKTKPEKFSLPKCSVQVGIKISSVHKRLASEKRENPVKKVM 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ATMRFLSSGKEQKTKPKEKVKTKPEKFSLPKCSVQVGIKISSVHKRLASEKRENPVKKVM 480

Qy        481 LASRSETSGKEAACESSTPSWASDHNYNAVKPEKPEKPTALSPTLLSKCTYHPKAGFPGP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LASRSETSGKEAACESSTPSWASDHNYNAVKPEKPEKPTALSPTLLSKCTYHPKAGFPGP 540

Qy        541 SHHLGGCLGLSRTRVLGVLVLIVASSSLPARSRYQDASGPQVFLPSLWSLSGWFLKSCVG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 SHHLGGCLGLSRTRVLGVLVLIVASSSLPARSRYQDASGPQVFLPSLWSLSGWFLKSCVG 600

Qy        601 LMLEAISYFSFRPW 614
              ||||||||||||||
Db        601 LMLEAISYFSFRPW 614


RESULT 2
AAB93638
ID   AAB93638 standard; protein; 562 AA.
XX
AC   AAB93638;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:13130.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
XX
PR   29-JUL-1999;   99JP-00248036.
PR   27-AUG-1999;   99JP-00300253.
PR   11-JAN-2000; 2000JP-00118776.
PR   02-MAY-2000; 2000JP-00183767.
PR   09-JUN-2000; 2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 13130; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 562 AA;

  Query Match             65.4%;  Score 2110;  DB 4;  Length 562;
  Best Local Similarity   75.0%;  Pred. No. 5.2e-138;
  Matches  420;  Conservative   32;  Mismatches   94;  Indels   14;  Gaps    7;

Qy          1 MDDKGHLSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPSEQQP--QQHNLS 58
              |||||  |||||||||||||||||||||||||||||||||||| |||| |  |  ||  ||
Db          1 MDDKGDPSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDAEADPLEPPPPPQQQLGLS 60

Qy         59 LRRSGRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSE 118
              |||||||||||||||:|||  ||||::::|||||||| ||||   || ||||||||||:||
Db         61 LRRSGRQPKRTERVEQFLTIARRRGRRSMPVSLEDSGEPTSCPATDAETASEGSVESASE 120

Qy        119 IRSGPVSDSLG-KEHPASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLRRKREQEPVERS 177
              |||| | |   || |||||| |||:: :|||||||||||||||||||||||||||| ||
Db        121 TRSGPQSASTAVKERPASSEKVKGGDDHDDTSDSDSDGLTLKELQNRLRRKREQEPTERP 180

Qy        178 LRGSQNRLRKKRREEDSAET-GSVQIGSAEQDRPLCKQEPEASQGPVSQSETDDIENQLE 236
              |:| |:|||||||||  ||| ||    : |   | |||||  || |||:  || |::||
```

```
Db        181 LKGIQSRLRKKRREEGPAETVGSEASDTVEGVLP-SKQEPENDQGVVSQAGKDDRESKLE 239

Qy        237 GKATQGNTEENPREAGKPKPECEVYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGI 296

Db        240 GKAAQDIKDEEPEDLGRPKPECEGYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGI 299

Qy        297 SEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQDSGCRSVGADGTDCTSIGTVEQK 356

Db        300 SEARGRLLERNGEDYICPNCTILQVQDETHSETADQQEAKWRPGDADGTDCTSIGTIEQK 359

Qy        357 SGEDQGIKGRIEKAANPSGKKKLKIFQPVVEAPGAPKCIGPGCSSVAQPDSVYCSNDCIL 416

Db        360 SSEDQGIKGRIEKAANPSGKKKLKIFQPVIEAPGASKCIGPGCCHVAQPDSVYCSNDCIL 419

Qy        417 KHAAATMRFLSSGKEQKTKPKEKVKTKPEKFSLPKCSVQVGIKISSVHKRLASEKRENPV 476

Db        420 KHAAATMKPLSSGKEQKPKPKEKMKMKPEKPSLPKCGAQAGIKISSVHKRPAPEKKETTV 479

Qy        477 KK-VMLASRSETSGKEAACESSTPSWASDHNYNAVKPEKPEKPTALSPTLLSKCTYHPKA 535

Db        480 KKAVVVPARSEALGKEAACESSTPSWASDHNYNAVKPEKTAAP---SPSLLYKCMYHLGV 536

Qy        536 GFPGPSHHL-----GGCLGL 550

Db        537 GLLDPSRSFWIAIPWACPGL 556
```

```
RESULT 3
AAY67579
ID   AAY67579 standard; protein; 562 AA.
XX
AC   AAY67579;
XX
DT   19-JUN-2000  (first entry)
XX
DE   Human death inducer-obliterator 1 (DIO-1) polypeptide.
XX
KW   Death inducer-obliterator 1; DIO-1; cell death; cancer; tumour; human;
KW   autoimmune disease; cytostatic; immunosuppressive; antidiabetic;
KW   antirheumatic; antiinflammatory; antiproliferative.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          165. .172
FT                   /note= "NLS sequence"
FT   Region          185. .193
FT                   /note= "NLS sequence"
FT   Region          271. .288
FT                   /note= "zinc finger motif"
FT   Region          293. .320
FT                   /note= "zinc finger motif"
XX
PN   WO200015787-A1.
XX
PD   23-MAR-2000.
XX
PF   10-SEP-1999;   99WO-GB003019.
XX
PR   10-SEP-1998;   98SE-00003069.
PR   17-SEP-1998;   98US-0100873P.
XX
PA   (CNSJ ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.
PA   (BANN/) BANNERMAN D G.
XX
PI   Alonso CM,  Domingo DG,  Grandien A,  Leonardo E,  Martinez P;
XX
DR   WPI; 2000-271426/23.
DR   N-PSDB; AAZ90578.
XX
PT   New DNA encoding human and murine death inducer-obliterator 1
PT   polypeptides, useful in the treatment of cancer, autoimmune diseases,
PT   diabetes, rheumatoid arthritis, benign tumors, malignant tumors and
PT   hyperproliferative skin disorders.
XX
PS   Claim 6; Fig 1C; 27pp; English.
XX
CC   The invention provides nucleic acids encoding the human and murine death
CC   inducer-obliterator 1 (DIO-1) polypeptides. The polypeptides can be
CC   expressed by standard recombinant methodology. The DIO-1 polypeptides,
CC   agonists and antagonists are used as a medicament for treating diseases
CC   characterized by an alteration in cell death or by hyperproliferation,
CC   e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign
CC   tumours, malignant tumours or hyperproliferative skin disorders. They are
CC   also useful in the treatment of metabolic, proliferative or inflammatory
CC   conditions. The present sequence represents the human DIO-1 polypeptide
XX
SQ   Sequence 562 AA;
```

```
  Query Match           65.3%;  Score 2108;  DB 3;  Length 562;
  Best Local Similarity  75.0%;  Pred. No. 7.1e-138;
  Matches  420;  Conservative   32;  Mismatches   94;  Indels   14;  Gaps    7;

Qy          1 MDDKGHLSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPSEQQP--QQHNLS 58

Db          1 MDDKGDPSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDAEADPLEPPPPQQQLGLS 60

Qy         59 LRRSGRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSE 118

Db         61 LRRSGRQPKRTERVEQFLTIARRRGRRSMPVSLEDSGEPTSCPATDAETASEGSVESASE 120

Qy        119 IRSGPVSDSLG-KEHPASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLRRKREQEPVERS 177

Db        121 TRSGPQSASTAVKERPASSEKVKGGDDHDDTSDSDSDGLTLKELQNRLRRKREQEPTERP 180

Qy        178 LRGSQNRLRKKRREEDSAET-GSVQIGSAEQDRPLCKQEPEASQGPVSQSETDDIENQLE 236

Db        181 LKGIQSRLRKKRREEGPAETVGSEASDTVEGVLP-SKQEPENDQGVVSQAGKDDRESKLE 239

Qy        237 GKATQGNTEENPREAGKPKPECEVYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGI 296

Db        240 GKAAQDIKDEEPGDLGRPKPECEGYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGI 299

Qy        297 SEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQDSGCRSVGADGTDCTSIGTVEQK 356

Db        300 SEARGRLLERNGEDYICPNCTILQVQDETHSETADQQEAKWRPGDADGTDCTSIGTIEQK 359

Qy        357 SGEDQGIKGRIEKAANPSGKKKLKIFQPVVEAPGAPKCIGPGCSSVAQPDSVYCSNDCIL 416

Db        360 SSEDQGIKGRIEKAANPSGKKKLKIFQPVIEAPGASKCIGPGCCHVAQPDSVYCSNDCIL 419

Qy        417 KHAAATMRFLSSGKEQKTKPKEKVKTKPEKFSLPKCSVQVGIKISSVHKRLASEKRENPV 476

Db        420 KHAAATMKPLSSGKEQKPKPKEKMKMKPEKPSLPKCGAQAGIKISSVHKRPAPEKKETTV 479

Qy        477 KK-VMLASRSETSGKEAACESSTPSWASDHNYNAVKPEKPEKPTALSPTLLSKCTYHPKA 535

Db        480 KKAVVVPARSEALGKEAACESSTPSWASDHNYNAVKPEKTAAP---SPSLLYKCMYHLGV 536

Qy        536 GFPGPSHHL-----GGCLGL 550

Db        537 GLLDPSRSFWIAIPWACPGL 556
```

```
RESULT 4
AAM80219
ID   AAM80219 standard; protein; 1191 AA.
XX
AC   AAM80219;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human protein SEQ ID NO 3865.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
```

```
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US004098.
XX
PR   03-FEB-2000; 2000US-00496914.
PR   27-APR-2000; 2000US-00560875.
PR   20-JUN-2000; 2000US-00598075.
PR   19-JUL-2000; 2000US-00620325.
PR   01-SEP-2000; 2000US-00654936.
PR   15-SEP-2000; 2000US-00663561.
PR   20-OCT-2000; 2000US-00693325.
PR   30-NOV-2000; 2000US-00728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI   Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI   Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   N-PSDB; AAK53352.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT   in diagnosis and gene therapy.
XX
PS   Claim 20; Page 455-456; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC   (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC   sequence listing were missing at the time of publication
XX
SQ   Sequence 1191 AA;

  Query Match             55.5%;  Score 1792.5;  DB 4;  Length 1191;
  Best Local Similarity   70.6%;  Pred. No. 1.5e-115;
  Matches  358;  Conservative   32;  Mismatches   74;  Indels   43;  Gaps    6;

Qy         63 GRQPKRTERVEEPLTTVRRRGKKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSEIRSG 122
Db         22 GRQPKRTERVEQPLTIARRRGRRSMPVSLEDSGEPTSCPATDAETASEGSVESASETRSG 81

Qy        123 PVSDSLG-KEHPASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLRRKREQEPVERSLRGS 181
Db         82 PQSASTAVKERPASSEKVKGGDDHDDTSDSDSDGLTLKELQNRLRRKREQEPTERPLKGI 141

Qy        182 QNRLRKKRREEDSAET-GSVQIGSAEQDRPLCKQEPEASQGPVSQSETDDIENQLEGKAT 240
Db        142 QSRLRKKRREEGPAETVGSEASDTVEGVLP-SKQEPENDQGVVSQAGKDDRESKLEGKAA 200

Qy        241 QGNTEENPREAGKPKPECEVYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEAR 300
Db        201 QDIKDEEPGDLGRPKPECEGYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEAR 260

Qy        301 GRLLERNGEDYICPNCTILQVQDETNGSATNEQDSGCRSVGADGTDCTSIGTVEQKSGED 360
Db        261 GRLLERNGEDYICPNCTILQVQDETHSETADQQEAKWRPGDADGTDCTSIGTIEQKSSED 320

Qy        361 QGIKGRIEKAANPSGKKKLKIFQP---------------------------------- 384
Db        321 QGIKGRIEKAANPSGKKKLKIFQPGPGPGPVPTQLPVLWQVLEIAVSRSISAFTLLHCISCK 380

Qy        385 VVEAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAAATMRFLSSGKEQKTKPKEKVKTKP 444
Db        381 VIEAPGASKCIGPGCCHVAQPDSVYCSNDCILKHAAATMKFLSSGKEQKPKPKEKMKMKP 440

Qy        445 EKFSLPKCSVQVGIKISSVHKRLASEKRENPVKK-VMLASRSETSGKEAACESSTPSWAS 503
Db        441 EKPSLPKCGAQAGIKISSVHKRPAPEKKETTVKKAVVVPARSEALGKEAACESSTPSWAS 500

Qy        504 DHNYNAVKPEKPEKPTALSPTLLSKCT 530
Db        501 DHNYNAVKPEKTAAP---SPSLLYKST 524


RESULT 5
ABG22389
ID   ABG22389 standard; protein; 775 AA.
XX
AC   ABG22389;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #22380.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS86576.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 20; SEQ ID NO 52748; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
```

CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 775 AA;

Query Match 54.4%; Score 1756.5; DB 4; Length 775;
Best Local Similarity 57.1%; Pred. No. 2.7e-113;
Matches 378; Conservative 45; Mismatches 126; Indels 113; Gaps 14;

```
Qy          1 MDDKGHLSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEADPSEQQP--QQHNLS 58
              |||||  |||||||||||||||||||||||||||||||||||| |||| |  |  ||  ||
Db        131 MDDKGDPSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDAEADPLEPPPPQQQLGLS 190

Qy         59 LRRSGRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSE 118
              |||||||||||||||:|||  ||||::::|||||||| ||||   || ||||||||||:||
Db        191 LRRSGRQPKRTERVEQFLTIARRRGRRSMPVSLEDSGEPTSCPATDAETASEGSVESASE 250

Qy        119 IRSGPVSDSLG-KEHPASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLRRKREQEPVERS 177
              |||| | |   || |||||| |||:: :||||||||||||||||||||||||||||| ||
Db        251 TRSGPQSASTAVKERPASSEKVKGGDDHDDTSDSDSDGLTLKELQNRLRRKREQEPTERP 310

Qy        178 LRGSQNRLRKKRREEDSAET-GSVQIGSAEQDRPLCKQEPEASQGPVSQSETDDIENQLE 236
              |:| |:|||||||||  ||| ||    : |   |  |||||  || |||:  || |::||
Db        311 LKGIQSRLRKKRREEGPAETVGSEASDTVEGVLP-SKQEPENDQGVVSQAGKDDRESKLE 369

Qy        237 GKATQGNTEENPREAGKPKPECEVYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGI 296
              ||| |   :| | : |:|||||| ||||||||||||||||||||||||||||||||||||
Db        370 GKAAQDIKDEEPGDLGRPKPECEGYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGI 429

Qy        297 SEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQDSGCRSVG-ADGTDCTSIGTVEQ 355
              |||||||||||||||||||||||||||:    ::|::   :| ||||| |||| :
Db        430 SEARGRLLERNGEDYICPNCTILQVQDETHSETADQQEAXMGDLGDADGTDFTSIGNIRS 489

Qy        356 K--SGEDQGIKGRIEKAANPSGKKKLKIFQP----------------------VVEAPGA 391
              :  |   :|:|||||||||||||:||||||                      |:|||||
Db        490 RSLSRRPKGLKGRIEKAANPSGKKELKIFQPVLEIAVSRSISAFTLLHCISCKVIEAPGA 549

Qy        392 PKCIGPGCSSVAQPDSVYCSNDCILKHAAATMRFLSSGKEQKTKPKEKVKTKPEKFSLPK 451
              |||||||  |||||||||||||||||||||:|||||||| |||||:| |||| ||||
Db        550 SKCIGPGCCHVAQPDSVYCSNDCILKHAAATMKPLSSGKEQKPKPKEKMKMKPEKPSLPK 609

Qy        452 C------------SVQVG-------------------IKISSVHKRLASEKRENPVKKVM 480
              |            | |:|                   :| | :|   |   | : ::|:
Db        610 CGAQPVQWLLLLSSWQLGTSSGLQVTLCQPPPTEHKWLKFSVLHGETAGPPRAHRIRKL- 668

Qy        481 LASRSETSGKEAACESSTPSWASDHNYNAVKPEKPEKPTALSPTLLSKCTYHPKAGPPGP 540
                          | : |:   ::|           |  |          |:|
Db        669 -----------------NKYWGACHHCMPLRP---------LPCL---------PGWPRS 693

Qy        541 SHHLGGCLGLSRTRV-------LGVLVLIVASSSLPARSRYQDASGPQVFLPSLWSLSGW 593
              |  | |   | | |       | :          |   |   | ||       |    |
Db        694 SPSLVGAGSLPRHAVHLNEPCPLAI---------CPLARRLSPARGPLGSDDLTWGYKVW 744

Qy        594 FL 595
              ||
Db        745 LL 746
```

RESULT 6
AAM79235
ID AAM79235 standard; protein; 647 AA.
XX
AC AAM79235;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 1897.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US004098.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR N-PSDB; AAK52368.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.
XX
PS Claim 20; Page 4293-4294; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
SQ Sequence 647 AA;

Query Match 52.5%; Score 1693.5; DB 4; Length 647;
Best Local Similarity 70.4%; Pred. No. 5.1e-109;
Matches 340; Conservative 28; Mismatches 72; Indels 43; Gaps 6;

```
Qy         87 VPVSLEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDSLG-KEHPASSEKAKGGEEE 145
              :|||||| ||||   || ||||||||||:|| |||| | |   || |||||| |||::
Db          1 MPVSLEDSGEPTSCPATDAETASEGSVESASETRSGPQSASTAVKERPASSEKVKGGDDH 60

Qy        146 EDTSDSDSDGLTLKELQNRLRRKREQEPVERSLRGSQNRLRKKRREEDSAET-GSVQIGS 204
              :|||||||||||||||||||||||||||||| || |:| |:|||||||||  ||| ||    :
Db         61 DDTSDSDSDGLTLKELQNRLRRKREQEPTERPLKGIQSRLRKKRREEGPAETVGSEASDT 120

Qy        205 AEQDRPLCKQEPEASQGPVSQSETDDIENQLEGKATQGNTEENPREAGKPKPECEVYDPN 264
              |   |  ||||| || |||:  || |::||||| |   :| | : |:|||||| ||||
Db        121 VEGVLP-SKQEPENDQGVVSQAGKDDRESKLEGKAAQDIKDEEPGDLGRPKPECEGYDPN 179

Qy        265 ALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDE 324
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 ALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDE 239

Qy        325 TNGSATNEQDSGCRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLKIFQP 384
              |:    ::|::  |  ||||||||||:|||| ||||||||||||||||||||||||||
Db        240 THSETADQQEAKWRPGDADGTDCTSIGTIEQKSSEDQGIKGRIEKAANPSGKKKLKIFQP 299
```

```
Qy        385 ----------------------------------VVEAPGAPKCIGPGCSSVAQPDSV 408
                                                |:||||| |||||||  |||||||
Db        300 GPGPVPTQLPVLWQVLEIAVSRSISAFTLLHCISCKVIEAPGASKCIGPGCCHVAQPDSV 359

Qy        409 YCSNDCILKHAAATMRPLSSGKEQKTKPKEKVKTKPEKPSLPKCSVQVGIKISSVHKRLA 468
              |||||||||||||||:|||||||| |||||:| |||| |||||  | |||||||||| |
Db        360 YCSNDCILKHAAATMKPLSSGKEQKPKPKEKMKMKPEKPSLPKCGAQAGIKISSVHKRPA 419

Qy        469 SEKRENPVKK-VMLASRSETSGKEAACESSTPSWASDHNYNAVKPEKPEKPTALSPTLLS 527
                ||:|  ||| |:: :|||  |||||||||||||||||||||||||   |   ||:||
Db        420 PEKKETTVKKAVVVPARSEALGKEAACESSTPSWASDHNYNAVKPEKTAAP---SPSLLY 476

Qy        528 KCT 530
                | |
Db        477 KST 479
```

RESULT 7
ABG22388
ID ABG22388 standard; protein; 330 AA.
XX
AC ABG22388;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #22379.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS86575.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 52747; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 330 AA;

```
  Query Match             37.9%;  Score 1224.5;  DB 4;  Length 330;
  Best Local Similarity   70.7%;  Pred. No. 7.9e-77;
  Matches  232;  Conservative   21;  Mismatches   38;  Indels   37;  Gaps    2;

Qy        232 ENQLEGKATQGNTEENPREAGKPKPECEVYDPNALYCICRQPHNNRFMICCDRCEEWFHG 291
              |::||||| |   :| | : |:|||||| |||||||||||||||||||||||||||||||
Db          1 ESKLEGKAAQDIKDEEPGDLGRPKPECEGYDPNALYCICRQPHNNRFMICCDRCEEWFHG 60

Qy        292 DCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQDSGCRSVGADGTDCTSIG 351
              ||||||||||||||||||||||||||||||:    ::|::  |   ||||||||||
Db         61 DCVGISEARGRLLERNGEDYICPNCTILQVQDETHSETADQQEAKWRPGDADGTDCTSIG 120

Qy        352 TVEQKSGEDQGIKGRIEKAANPSGKKKLKIFQP---------------------------- 384
              |:|||| ||||||||||||||||||||||||||
Db        121 TIEQKSSEDQGIKGRIEKAANPSGKKKLKIFQPGPGPGPVPTQLPVLWQVLEIAVSRSISAF 180

Qy        385 ---------VVEAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAAATMRPLSSGKEQKTK 435
                       |:||||| |||||||  |||||||||||||||||||||:||||||||| |
Db        181 TLLHCISCKVIEAPGASKCIGPGCCHVAQPDSVYCSNDCILKHAAATMKPLSSGKEQKPK 240

Qy        436 PKEKVKTKPEKPSLPKCSVQVGIKISSVHKRLASEKRENPVKK-VMLASRSETSGKEAAC 494
              ||||:| |||| |||||  | |||||||||| | ||:|  ||| |:: :|||  ||||||
Db        241 PKEKMKMKPEKPSLPKCGAQAGIKISSVHKRPAPEKKETTVKKAVVVPARSEALGKEAAC 300

Qy        495 ESSTPSWASDHNYNAVKPEKPEKPTALS 522
              |||||||||||||||||||   |: ::
Db        301 ESSTPSWASDHNYNAVKPEKTAAPSPVT 328
```

RESULT 8
AAB43724
ID AAB43724 standard; protein; 181 AA.
XX
AC AAB43724;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human cancer associated protein sequence SEQ ID NO:1169.
XX
KW Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnerary; immunomodulator;
KW antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening.
XX
OS Homo sapiens.
XX
PN WO200055350-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US005882.
XX
PR 12-MAR-1999; 99US-0124270P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587533/55.
DR N-PSDB; AAC77933.
XX

```
PT   Novel isolated nucleic acids comprising sequences encoding peptides
PT   useful for treating or diagnosing e.g. cancer.
XX
PS   Claim 11; Page 1790-1791; 2352pp; English.
XX
CC   AAC77607 to AAC78448 encode the human cancer associated proteins given in
CC   AAB43398 to AAB44239. The proteins can have activities based on the
CC   tissues and cells the genes are expressed in. Example of activities
CC   include: cytostatic; proliferative; vulnerary; immunomodulator;
CC   antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC   antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC   dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC   nootropic; vasotropic; antipsoriatic and antiangiogenic. The
CC   polynucleotides and polypeptides can be used for preventing, treating or
CC   ameliorating medical conditions and diagnosing pathological conditions.
CC   Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC   the present invention may be used to treat immune disorders by activating
CC   or inhibiting the proliferation, differentiation or mobilisation of
CC   immune cells, to treat disorders of haematopoietic cells, autoimmune
CC   disorders, allergic reactions, graft versus host disease and organ
CC   rejection, modulate haemostatic or thrombolytic activity, modulate
CC   inflammation, cancers, cardiovascular disorders, neurological disease and
CC   bacterial or viral infections. The peptides, nucleotides, antibodies,
CC   agonists and antagonists may be also be used in drug screens. AAC78449 to
CC   AAC78457 and AAB44240 represent sequences used in the exemplification of
CC   the present invention
XX
SQ   Sequence 181 AA;

  Query Match             22.9%;  Score 740;  DB 3;  Length 181;
  Best Local Similarity   85.1%;  Pred. No. 1.6e-43;
  Matches  131;  Conservative    8;  Mismatches   15;  Indels    0;  Gaps    0;

Qy        252 GKPKPECEVYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDY 311
              |:|||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db         11 GRPKPECEGYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDY 70

Qy        312 ICPNCTILQVQDETNGSATNEQDSGCRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAA 371
              |||||||||||||:    ::|::  |   ||||||||||:|||| ||||||||||||
Db         71 ICPNCTILQVQDETHSETADQQEAKWRPGDADGTDCTSIGTIEQKSSEDQGIKGRIEKAA 130

Qy        372 NPSGKKKLKIFQPVVEAPGAPKCIGPGCSSVAQP 405
              |||||||||||||:||||| |||||||  || |
Db        131 NPSGKKKLKIFQPVIEAPGASKCIGPGCCHVAHP 164


RESULT 9
ABG22387
ID   ABG22387 standard; protein; 167 AA.
XX
AC   ABG22387;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #22378.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS86574.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 20; SEQ ID NO 52746; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC   amino acid sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 167 AA;

  Query Match             22.7%;  Score 732;  DB 4;  Length 167;
  Best Local Similarity   83.1%;  Pred. No. 5.2e-43;
  Matches  138;  Conservative    9;  Mismatches   19;  Indels    0;  Gaps    0;

Qy        301 GRLLERNGEDYICPNCTILQVQDETNGSATNEQDSGCRSVGADGTDCTSIGTVEQKSGED 360
              |||||||||||||||||||||||:    ::|::  |   ||||||||||:|||| ||
Db          1 GRLLERNGEDYICPNCTILQVQDETHSETADQQEAKWRPGDADGTDCTSIGTIEQKSSED 60

Qy        361 QGIKGRIEKAANPSGKKKLKIFQPVVEAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAA 420
              |||||||||||||||||||||||:|||||  |||||||  ||||||||||||||||||||
Db         61 QGIKGRIEKAANPSGKKKLKIFQPVIEAPGASKCIGPGCCHVAQPDSVYCSNDCILKHAA 120

Qy        421 ATMRPLSSGKEQKTKPKEKVKTKPEKPSLPKCSVQVGIKISSVHKR 466
              |||:|||||||| ||||:| |||| |||||  | ||||||||||
Db        121 ATMKPLSSGKEQKPKPKEKMKMKPEKPSLPKCGAQAGIKISSVHKR 166


RESULT 10
ABP05496
ID   ABP05496 standard; protein; 108 AA.
XX
AC   ABP05496;
XX
DT   24-JUN-2002  (first entry)
XX
DE   Human ORFX protein sequence SEQ ID NO:10974.
XX
KW   Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW   hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW   degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW   cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW   hypertension; hypothyroidism; cholesterol ester storage disease;
KW   immune deficiency; immune disorder; infectious disease;
KW   autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW   myasthenia gravis.
XX
OS   Homo sapiens.
XX
```

PN WO200192523-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US010836.
XX
PR 30-MAY-2000; 2000US-0206132P.
PR 29-AUG-2000; 2000US-0228716P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach MD;
XX
DR WPI; 2002-106308/14.
DR N-PSDB; ABN21248.
XX
PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders.
XX
PS Disclosure; SEQ ID NO 10974; 1037pp; English.
XX
CC The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage. N.B. The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 108 AA;

Query Match 17.3%; Score 559; DB 5; Length 108;
Best Local Similarity 99.1%; Pred. No. 3.1e-31;
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy        357 SGEDQGIKGRIEKAANPSGKKKLKIFQPVVEAPGAPKCIGPGCSSVAQPDSVYCSNDCIL 416
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 SGEDHGIKGRIEKAANPSGKKKLKIFQPVVEAPGAPKCIGPGCSSVAQPDSVYCSNDCIL 60

Qy        417 KHAAATMRFLSSGKEQKTKPKEKVKTKPEKPSLPKCSVQVGIKISSVH 464
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 KHAAATMRFLSSGKEQKTKPKEKVKTKPEKPSLPKCSVQVGIKISSVH 108
```

RESULT 11
ABB71487
ID ABB71487 standard; protein; 2016 AA.
XX
AC ABB71487;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 41253.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL15590.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 41253; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2016 AA;

Query Match 10.0%; Score 323.5; DB 4; Length 2016;
Best Local Similarity 22.7%; Pred. No. 2.7e-13;
Matches 160; Conservative 86; Mismatches 253; Indels 205; Gaps 30;

```
Qy          8 SNEEAPKAIK---PTSKEFRKTWGFRRTTI---AKREGAGD-TEADPSEQQPQ------- 53
              ::|: || :|    |: :         | |    : :| |:   |||:
Db        639 ADEDIPKLVKMLESTAADLDPPPVSLPTNIFPAPEMQGVGNAAAADPNNIMDTANEDEIT 698

Qy         54 ----QHNLSLRRSGRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTVTDVETAS 109
                 || : |    :|    | |:: | :   |  :  ||:  | ||
Db        699 ADFLQHVVGLIEEDKQ-FEAEVVKQVLAST-EPGTLDAIVSMPTSIEPVDVPQAHTNLLP 756

Qy        110 EGSV-ESSSEIRSGPV-----SDSLGKEHPASSEKAKGGEEEEDTSDSDSDGLTLKELQN 163
                |: | :  : | |:     | |:    | :| |   |             : |  :
Db        757 NASLTEPAQSMTSLPIACSTPSRSVAASTPPTSAKVVRGYGRV---------IYLPPI-- 805

Qy        164 RLRRKREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSAEQD----RPLCKQEPEAS 219
                : |  |: | :|           |:: |:  | :  |    :||
Db        806 -------EAPTTRAKRRAQFPSAPGMAATSSSDAGNLSPGESSLDASINQPL-------- 850

Qy        220 QGPVSQSETDDIENQLEGKATQGNTEENPREAGKPK-----------------PEC--- 258
                  | : |  : :   |    ||||    :                      |:
Db        851 -------NTSSLSN--DSQPGSGPKRPNPREPSMARRSTAPRRSKKLDASQNNDPDASES 901

Qy        259 --EVYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNC 316
                :  ||| |:|||||||||||||||| ||:|||| |||:::| |  :|  | |: || |
Db        902 QEDDDDPNKLWCICRQPHNNRFMICCDLCEDWFHGTCVGVTKAMGTDMENKGIDWKCPKC 961

Qy        317 TILQVQDETNGSATNEQDSGCRSVGADGTDCTSIGTVEQKSGEDQGI--KGRIEKAANPS 374
                |          |: | |       | |:    ||:  | : :     | : | ||
Db        962 VKRQ----------EERSQPRITDMLVTRPTT--QPEQRPSETKVLTTTAEIVQVAAPS 1008

Qy        375 G-KKKLKIFQPVVEAP-------------GAPK----------CIGPG----------- 398
                :: | :   |  :|             ||             :||
```

```
Db       1009 APRRTLPVVLTVASSPMRIPMAKPAKKFPTGAISHQQQQQLNFIRLGPSPGKRISETLCV 1068

Qy        399 -CSSVAQPDSVYCSNDCILKHAAATMRFLSSGK-----------------EQKTKPKEK 439

Db       1069 VCKRPASTSSVYCGEECIRKYAQSAIQAHAATKGPLPQNAGAQSLLNNSFDAKKNKKKDL 1128

Qy        440 VK---------TKPEK---FSLPKCSVQVGIKISSVHKRLASEKRENPVKKVMLASRSET 487

Db       1129 FEDVLRQADTVSKVERINVFERKSGRVITGHMAPSAH-QFRKWLQENPSFEVLPSGTVQS 1187

Qy        488 SGKEAACESSTPSWASDHNYNAV-----KPEKPEKPTALSPTLLSKCTYHPKAGPPGPSH 542

Db       1188 ADAEKRLLKGAPEAATSTSEPAVLGVAKKP--PEGPAKLS---------HPQNTTVQASH 1236

Qy        543 HLGGCLGLSRTRVLGVLVLIVASSSLPARSRYQDASGPQVFLPS 586

Db       1237 QLG-----------------ISSVRPLAKKDKEKTTPTVQAPT 1262
```

RESULT 12
AAM39234
ID AAM39234 standard; protein; 1728 AA.
XX
AC AAM39234;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 2379.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI58390.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
PS Example 4; SEQ ID NO 2379; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 1728 AA;

Query Match 7.4%; Score 239.5; DB 4; Length 1728;
Best Local Similarity 20.5%; Pred. No. 1.5e-07;
Matches 101; Conservative 75; Mismatches 163; Indels 153; Gaps 17;

```
Qy         90 SLEDSSEPTSSTVTDVETASEGSVESSSEIRS--GPVSDSLGK--------EHPASSEKA 139

Db        155 NLQDDRNSQSSSVSYLESKSVKSKHTKPVIHSKQNMTTDAPKKIVAAKYEVIHSKTKVNV 214

Qy        140 KGGEEEEDTSDSDSDGLTLKELQN-----RLRRKR-EQEPV-------ERSLRGSQNRLR 186

Db        215 KSVKRNTDVPESQ---------QNFHRPVKVRKKQIDKEPKIQSCNSGVKSVKNQAHSVL 265

Qy        187 KKRREEDSAETGSVQI--------GSAEQDRPLCKQEPE--ASQGPVSQSE--------- 227

Db        266 KKTLQDQTL----VQIFKPLTHSLSDKSHAHPGCLKEPHHPAQTGHVSHSSQKQCHKPQQ 321

Qy        228 -------TDDIENQLEGKATQGNTEENPREAGKPKPECE---------------------- 259

Db        322 QAPAMKTNSHVKEELEHPGVEHFKEEDKLKLKKPEKNLQPRQRRSSKSFSLDEPPLFIPD 381

Qy        260 --------------------VYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEA 299

Db        382 NIATIRREGSDHSSSFESKYMWTPSKQCGFCKKPHGNRFMVGCGRCDDWFHGDCVGLSLS 441

Qy        300 RGRLLERNGEDYICPNC-------TILQVQDETNGSATNEQDSGCRSVGADGTDCTSIGT 352

Db        442 QAQQMGEEDKEYVCVKCCAEEDKKTEILDPDTLENQATVEPHSGDKTM-----ECEKLGL 496

Qy        353 VEQKSGE-----DQGIKGRIEKAANPSGKKKLKIFQPVVEAPGAPKCIGPGCSSVAQPDS 407

Db        497 SKHTTNDRTKYIDDTVKHKVKILKRESGEGR--------------------------NS 529

Qy        408 VYCSNDCILKHAAATMRFLSSGKEQKTKPKEKVKTKPEKPSLPKCSVQVGIKISSVHKRL 467

Db        530 SDCRDNEIKKWQLAPLRKMG----QPVLPRRSSEEKSEK--IPKESTTVTCTGEKASKPG 583

Qy        468 ASEKRENPVKKV 479

Db        584 THEKQEMKKKKV 595
```

RESULT 13
AAM79892
ID AAM79892 standard; protein; 1766 AA.
XX
AC AAM79892;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 3538.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX

PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US004098.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR N-PSDB; AAK53025.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.
XX
PS Claim 20; Page 376-377; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
SQ Sequence 1766 AA;

Query Match 7.4%; Score 239.5; DB 4; Length 1766;
Best Local Similarity 20.5%; Pred. No. 1.6e-07;
Matches 101; Conservative 75; Mismatches 163; Indels 153; Gaps 17;

```
Qy         90 SLEDSSEPTSSTVTDVETASEGSVESSSEIRS--GPVSDSLGK--------EHPASSEKA 139
              :|:|     ||:|: :|: |  |  :   | |     :|:  |          |  :
Db        193 NLQDDRNSQSSSVSYLESKSVKSKHTKPVIHSKQNMTTDAPKKIVAAKYEVIHSKTKVNV 252

Qy        140 KGGEEEEDTSDSDSDGLTLKELQN-----RLRRKR-EQEPV-------ERSLRGSQNRLR 186
              |  :   | :|          ||      ::|:|: ::||          :|::   : :
Db        253 KSVKRNTDVPESQ---------QNFHRPVKVRKKQIDKEPKIQSCNSGVKSVKNQAHSVL 303

Qy        187 KKRREEDSAETGSVQI--------GSAEQDRPLCKQEPE--ASQGPVSQSE--------- 227
              ||  :: :     |||                | | :||   |  | || |
Db        304 KKTLQDQTL----VQIFKPLTHSLSDKSHAHPGCLKEPHHPAQTGHVSHSSQKQCHKPQQ 359

Qy        228 -------TDDIENQLEGKATQGNTEENPREAGKPKPECE--------------------- 259
                    :: :||     :   ||:  :  ||:   :
Db        360 QAPAMKTNSHVKEELEHPGVEHFKEEDKLKLKKPEKNLQPRQRRSSKSFSLDEPPLFIPD 419

Qy        260 -------------------VYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEA 299
                                :: |:     |::|| |||| | ||::|||||||:| :
Db        420 NIATIRREGSDHSSSFESKYMWTPSKQCGFCKKPHGNRFMVGCGRCDDWFHGDCVGLSLS 479

Qy        300 RGRLLERNGEDYICPNC-------TILQVQDETNGSATNEQDSGCRSVGADGTDCTSIGT 352
              : : :    ::|:|  |       | :    |     || |  || ::::     :|  :|
Db        480 QAQQMGEEDKEYVCVKCCAEEDKKTEILDPDTLENQATVEPHSGDKTM-----ECEKLGL 534

Qy        353 VEQKSGE-----DQGIKGRIEKAANPSGKKKLKIFQPVVEAPGAPKCIGPGCSSVAQPDS 407
                  : : :     |  :| :::     ||: :                     :|
```

Db 535 SKHTTNDRTKYIDDTVKHKVKILKRESGEGR----------NS 567

Qy 408 VYCSNDCILKHAAATMRPLSSGKEQKTKPKEKVKTKPEKPSLPKCSVQVGIKISSVHKRL 467
| :: | | | :| : | |: : | || :|| | | |
Db 568 SDCRDNEIKKWQLAPLRKMG----QPVLPRRSSEEKSEK--IPKESTTVTCTGEKASKPG 621

Qy 468 ASEKRENPVKKV 479
||:| |||
Db 622 THEKQEMKKKKV 633

RESULT 14
AAM41020
ID AAM41020 standard; protein; 1766 AA.
XX
AC AAM41020;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 5951.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI60176.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
PS Example 2; SEQ ID NO 5951; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and

```
CC   C.N.S disorders. Note: The sequence data for this patent did not form
CC   part of the printed specification
XX
SQ   Sequence 1766 AA;

  Query Match             7.4%;  Score 239.5;  DB 4;  Length 1766;
  Best Local Similarity   20.5%;  Pred. No. 1.6e-07;
  Matches  101;  Conservative   75;  Mismatches  163;  Indels  153;  Gaps   17;

Qy         90 SLEDSSEPTSSTVTDVETASEGSVESSSEIRS--GPVSDSLGK--------EHPASSEKA 139
              :|:|     ||:|: :|: |  |  :   | |     :|:  |         |  :
Db        193 NLQDDRNSQSSSVSYLESKSVKSKHTKPVIHSKQNMTTDAPKKIVAAKYEVIHSKTKVNV 252

Qy        140 KGGEEEEDTSDSDSDGLTLKELQN-----RLRRKR-EQEPV-------ERSLRGSQNRLR 186
              |  :   | :|          ||     ::|:|: ::||         :|::   : :
Db        253 KSVKRNTDVPESQ---------QNFHRPVKVRKKQIDKEPKIQSCNSGVKSVKNQAHSVL 303

Qy        187 KKRREEDSAETGSVQI--------GSAEQDRPLCKQEPE--ASQGPVSQSE--------- 227
              ||  :: :     |||              | | :||    |  | || |
Db        304 KKTLQDQTL----VQIFKPLTHSLSDKSHAHPGCLKEPHHPAQTGHVSHSSQKQCHKPQQ 359

Qy        228 -------TDDIENQLEGKATQGNTEENPREAGKPKPECE--------------------- 259
                    :: :||    :   ||:  :  ||:   :
Db        360 QAPAMKTNSHVKEELEHPGVEHFKEEDKLKLKKPEKNLQPRQRRSSKSFSLDEPPLFIPD 419

Qy        260 --------------------VYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEA 299
                                 :: |:     |::|| |||| | ||::|||||||:| :
Db        420 NIATIRREGSDHSSSFESKYMWTPSKQCGFCKKPHGNRFMVGCGRCDDWFHGDCVGLSLS 479

Qy        300 RGRLLERNGEDYICPNC-------TILQVQDETNGSATNEQDSGCRSVGADGTDCTSIGT 352
              : : :    ::|:|  |         | :   |     || |  || :::     :|  :|
Db        480 QAQQMGEEDKEYVCVKCCAEEDKKTEILDPDTLENQATVEFHSGDKTM-----ECEKLGL 534

Qy        353 VEQKSGE-----DQGIKGRIEKAANPSGKKKLKIFQPVVEAPGAPKCIGPGCSSVAQPDS 407
               :  : :      |  :| :::     ||: :                        :|
Db        535 SKHTTNDRTKYIDDTVKHKVKILKRESGEGR--------------------------NS 567

Qy        408 VYCSNDCILKHAAATMRFLSSGKEQKTKPKEKVKTKPEKFSLPKCSVQVGIKISSVHKRL 467
              | :: | |   | :| :      |   |:   : | ||  :|| |  |         |
Db        568 SDCRDNEIKKWQLAPLRKMG----QPVLPRRSSEEKSEK--IPKESTTVTCTGEKASKPG 621

Qy        468 ASEKRENPVKKV 479
                ||:|   |||
Db        622 THEKQEMKKKKV 633


RESULT 15
ABR64258
ID   ABR64258 standard; protein; 2039 AA.
XX
AC   ABR64258;
XX
DT   15-OCT-2003  (first entry)
XX
DE   Angiogenesis protein BNO168.
XX
KW   Cytostatic; antirheumatic; antiarthritic; antidiabetic; ophthalmological;
KW   antipsoriatic; antiarteriosclerotic; cardiant; vasotropic; angiogenesis;
KW   gene therapy; vasculature; cancer; rheumatoid arthritis; psoriasis;
KW   diabetic retinopathy; cardiovascular disease; atherosclerosis;
KW   ischemic limb disease; coronary artery disease.
XX
OS   Homo sapiens.
XX
PN   WO2003027285-A1.
XX
PD   03-APR-2003.
XX
PF   19-SEP-2002; 2002WO-AU001282.
XX
PR   27-SEP-2001; 2001AU-00007973.
PR   27-SEP-2001; 2001AU-00007974.
PR   11-OCT-2001; 2001AU-00008210.
PR   29-OCT-2001; 2001AU-00008532.
PR   13-NOV-2001; 2001AU-00008838.
PR   28-AUG-2002; 2002AU-00951032.
XX
PA   (BION-) BIONOMICS LTD.
XX
PI   Gamble JR,  Hahn CN,  Vadas MA;
XX
DR   WPI; 2003-354655/33.
DR   N-PSDB; ACF34533.
XX
PT   New angiogenic genes and polypeptides, useful for diagnosing,
PT   prognosticating or treating an angiogenesis-related disorder, e.g.
PT   cancer, rheumatoid arthritis, diabetic retinopathy, psoriasis or
PT   cardiovascular diseases.
XX
PS   Claim 15; SEQ ID NO 193; 90pp; English.
XX
CC   The invention relates to the isolation of novel genes (ACF34446-ACF34559)
CC   encoding proteins (ABR64180-ABR64281) involved in the process of
CC   angiogenesis. The nucleic acid molecules are useful in identifying and/or
CC   obtaining full-length human genes involved in an angiogenic process. The
CC   nucleic acid molecule, polypeptides or complexes encoded, cells or
CC   genetically modified non-human animals derived from these are useful for
CC   the screening of candidate pharmaceutical compounds used in treating
CC   angiogenesis-related disorders. They are also useful for diagnosing,
CC   prognosticating or treating an angiogenesis-related disorder, which
CC   involves uncontrolled or enhanced angiogenesis or is a disorder in which
CC   a decreased vasculature is of benefit (e.g. cancer, rheumatoid arthritis,
CC   diabetic retinopathy, psoriasis or cardiovascular diseases such as
CC   atherosclerosis), or involves inappropriately arrested or decreased
CC   angiogenesis or is a disorder in which an expanding vasculature is of
CC   benefit (e.g. ischemic limb disease or coronary artery disease). The
CC   modulator of expression or activity of the polypeptide encoded by the
CC   nucleic acid sequence is useful for manufacturing a medicament for the
CC   treatment of an angiogenesis-related disorder. This sequence corresponds
CC   to one of the novel angiogenic protein
XX
SQ   Sequence 2039 AA;

  Query Match             7.4%;  Score 239.5;  DB 6;  Length 2039;
  Best Local Similarity   20.5%;  Pred. No. 1.9e-07;
  Matches  101;  Conservative   75;  Mismatches  163;  Indels  153;  Gaps   17;

Qy         90 SLEDSSEPTSSTVTDVETASEGSVESSSEIRS--GPVSDSLGK--------EHPASSEKA 139
              :|:|     ||:|: :|: |  |  :   | |     :|:  |         |  :
Db        466 NLQDDRNSQSSSVSYLESKSVKSKHTKPVIHSKQNMTTDAPKKIVAAKYEVIHSKTKVNV 525

Qy        140 KGGEEEEDTSDSDSDGLTLKELQN-----RLRRKR-EQEPV-------ERSLRGSQNRLR 186
              |  :   | :|          ||     ::|:|: ::||         :|::   : :
Db        526 KSVKRNTDVPESQ---------QNFHRPVKVRKKQIDKEPKIQSCNSGVKSVKNQAHSVL 576

Qy        187 KKRREEDSAETGSVQI--------GSAEQDRPLCKQEPE--ASQGPVSQSE--------- 227
              ||  :: :     |||              | | :||    |  | || |
Db        577 KKTLQDQTL----VQIFKPLTHSLSDKSHAHPGCLKEPHHPAQTGHVSHSSQKQCHKPQQ 632

Qy        228 -------TDDIENQLEGKATQGNTEENPREAGKPKPECE--------------------- 259
                    :: :||    :   ||:  :  ||:   :
Db        633 QAPAMKTNSHVKEELEHPGVEHFKEEDKLKLKKPEKNLQPRQRRSSKSFSLDEPPLFIPD 692

Qy        260 --------------------VYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEA 299
                                 :: |:     |::|| |||| | ||::|||||||:| :
Db        693 NIATIRREGSDHSSSFESKYMWTPSKQCGFCKKPHGNRFMVGCGRCDDWFHGDCVGLSLS 752

Qy        300 RGRLLERNGEDYICPNC-------TILQVQDETNGSATNEQDSGCRSVGADGTDCTSIGT 352
              : : :    ::|:|  |         | :   |     || |  || :::     :|  :|
Db        753 QAQQMGEEDKEYVCVKCCAEEDKKTEILDPDTLENQATVEFHSGDKTM-----ECEKLGL 807

Qy        353 VEQKSGE-----DQGIKGRIEKAANPSGKKKLKIFQPVVEAPGAPKCIGPGCSSVAQPDS 407
               :  : :      |  :| :::     ||: :                        :|
Db        808 SKHTTNDRTKYIDDTVKHKVKILKRESGEGR--------------------------NS 840

Qy        408 VYCSNDCILKHAAATMRFLSSGKEQKTKPKEKVKTKPEKFSLPKCSVQVGIKISSVHKRL 467
```

```
                 | :: | |   | :| :     |   |:   : | ||  :|| |  |        |
Db        841 SDCRDNEIKKWQLAPLRKMG----QPVLPRRSSEEKSEK--IPKESTTVTCTGEKASKPG 894

Qy        468 ASEKRENPVKKV 479
                ||:|   |||
Db        895 THEKQEMKKKKV 906


Search completed: April 20, 2004, 21:18:22
Job time : 63 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 23, 2004, 07:20:53 ; Search time 7067.26 Seconds
(without alignments)
12114.300 Million cell updates/sec

Title: US-09-787-016A-2
Perfect score: 2867
Sequence: 1 tccgtggtagctctggaaat.....tttgtttccttttgtgctggg 2867

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2828.2	98.6	4619	11	AK042474	AK042474 Mus muscu
	2	2775.8	96.8	4651	11	BC060249	BC060249 Mus muscu
	3	1482.4	51.7	1614	11	AK014422	AK014422 Mus muscu
	4	1398.8	48.8	1640	11	BC029110	BC029110 Mus muscu
	5	1010.8	35.3	2752	11	BC014615	BC014615 Homo sapi
	6	835.6	29.1	3385	11	AK044919	AK044919 Mus muscu
	7	780.8	27.2	874	13	BU522670	BU522670 AGENCOURT
	8	702.2	24.5	796	12	BI078113	BI078113 602872420
	9	701	24.5	982	13	BY713603	BY713603 BY713603
	10	665.4	23.2	667	14	CF169555	CF169555 B0815C09-
	11	621.4	21.7	691	10	BB637982	BB637982 BB637982
	12	617.4	21.5	632	13	BQ771134	BQ771134 UI-M-FI0-
	13	607.4	21.2	623	14	CF535718	CF535718 UI-M-GH0-
	14	594.6	20.7	645	10	BB636983	BB636983 BB636983
	15	591.6	20.6	746	13	BY752619	BY752619 BY752619
	16	591.4	20.6	639	10	BB664208	BB664208 BB664208
	17	587.4	20.5	613	14	CB518822	CB518822 UI-M-GH0-
	18	579.4	20.2	693	10	BB292563	BB292563 BB292563
	19	570.6	19.9	690	14	CA327345	CA327345 UI-M-FY0-
c	20	516.8	18.0	1086	9	AL578222	AL578222 AL578222
	21	494	17.2	925	9	AL546121	AL546121 AL546121
	22	482.8	16.8	486	28	CC156300	CC156300 XB629 Bay
	23	481.4	16.8	492	12	BG342505	BG342505 602374063
	24	469.4	16.4	491	12	BG342512	BG342512 602374072
	25	467.2	16.3	670	12	BI653462	BI653462 603300427
	26	450.4	15.7	521	14	CF167818	CF167818 B0790E04-
	27	448.2	15.6	476	28	AZ390149	AZ390149 1M0151B22
c	28	438.4	15.3	891	13	BX350413	BX350413 BX350413
	29	428	14.9	653	10	AW373892	AW373892 QV0-BT053
	30	428	14.9	653	10	AW373929	AW373929 QV3-BT053
	31	427.6	14.9	1093	9	AL555810	AL555810 AL555810
c	32	424.6	14.8	1018	9	AL550675	AL550675 AL550675
	33	420.2	14.7	1033	9	AL580669	AL580669 AL580669
	34	415.2	14.5	495	9	AI549786	AI549786 ve78e11.y
	35	412.6	14.4	506	14	CD541350	CD541350 B0231A09-
	36	412.6	14.4	542	14	CD543468	CD543468 B0249C01-
	37	402	14.0	760	10	BE298017	BE298017 601118492
	38	398	13.9	773	12	BG545979	BG545979 602573278
	39	392	13.7	1201	9	AL514138	AL514138 AL514138
c	40	389.8	13.6	850	14	CD364822	CD364822 UI-H-FT2-
	41	389.2	13.6	772	10	BE299127	BE299127 601118879
	42	386.8	13.5	650	10	BE298593	BE298593 601119753
	43	386.4	13.5	868	10	BF306721	BF306721 601888796
	44	385.8	13.5	1119	9	AL550431	AL550431 AL550431
c	45	385.2	13.4	498	12	BM247400	BM247400 K0801D11-

ALIGNMENTS

RESULT 1
AK042474
LOCUS AK042474 4619 bp mRNA linear HTC 19-SEP-2003
DEFINITION Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched library, clone:A630095A07 product:death inducer-obliterator-1, full insert sequence.
ACCESSION AK042474
VERSION AK042474.1 GI:26335138
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

```
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 4619)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
            Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
            Trust/MRC building Addenbrookes Hospital Cambridge) whose
            assistance we gratefully acknowledge.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1. .4619
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:A630095A07"
                     /db_xref="MGI:2407519"
                     /db_xref="taxon:10090"
                     /clone="A630095A07"
                     /tissue_type="thymus"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="3 days neonate"
     CDS             299. .2143
                     /note="unnamed protein product; death
                     inducer-obliterator-1 (MGD|MGI:1344352, GB|AJ238332,
                     evidence: BLASTN, 99%, match=2599)
                     putative"
                     /codon_start=1
                     /protein_id="BAC31270.1"
                     /db_xref="GI:26335139"
                     /translation="MDDKGHLSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTE
                     VDPSEQQPQQHNLSLRRSGRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTV
                     TDVETASEGSVESSSEIRSGPVSDSLGKEHPASSEKAKGGEEEEDTSDSDSDGLTLKE
                     LQNRLRRKREQEPVERSLRGSQNRLRKKRREDSAETGSVQIGSAEQDRPLCKQEPEA
                     SQGPVSQSETDDIENQLEGKATQGNTEENPREAGKPKPECEVYDPNALYCICRQPHNN
                     RFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATDEQD
                     SGCRSVGADGTDCTSIGTIEQKSGEDQGIKGRIEKAANPSGKKKLKIFQPVVEAPGAP
                     KCIGPGCSSVAQPDSVYCSNDCILKHAAATMRPLSSGKEQKTKPKEKVKTKPEKPSLP
                     KCSVQVGIKISSVHKRLASEKRENPVKKVMLASRSETSGKEAACESSTPSWASDHNYN
                     AVKPEKPEKPTALSPTLLSKCTYHPKAGFPGPSHHLGGCLGLSRTRVLGVLVLIVASS
                     SLPARSRYQDASGPQVFLPSLWSLSGWFLKSCVGLMLEAISYFSFRPW"
ORIGIN

  Query Match             98.6%;  Score 2828.2;  DB 11;  Length 4619;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 2856;  Conservative    0;  Mismatches    8;  Indels    3;  Gaps    2;

Qy          1 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         68 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 127

Qy         61 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        128 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 187

Qy        121 ACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        188 ACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTT 247

Qy        181 TATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        248 TATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGAT 307

Qy        241 AAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        308 AAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTC 367

Qy        301 AGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        368 AGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACG 427

Qy        361 GAGGCGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGA 420
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        428 GAGGTGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGA 487

Qy        421 CGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        488 CGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGG 547

Qy        481 AAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        548 AAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACT 607

Qy        541 GATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        608 GATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCT 667

Qy        601 GTATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        668 GTATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAA 727

Qy        661 GAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAAC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        728 GAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAAC 787

Qy        721 CGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        788 CGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAAT 847
```

```
Qy        781 CGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGC 840
Db        848 CGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGC 907

Qy        841 AGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTG 900
Db        908 AGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTG 967

Qy        901 TCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAAT 960
Db        968 TCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAAT 1027

Qy        961 ACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCC 1020
Db       1028 ACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCC 1087

Qy       1021 AATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGAT 1080
Db       1088 AATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGAT 1147

Qy       1081 CGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTC 1140
Db       1148 CGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTC 1207

Qy       1141 CTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGAT 1200
Db       1208 CTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGAT 1267

Qy       1201 GAGACAAACGGTAGCGCCACCAATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGAT 1260
Db       1268 GAGACAAACGGTAGCGCCACCGATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGAT 1327

Qy       1261 GGCACAGACTGCACAAGCATAGGGACAGTAGAGCAGAAGTCCGGAGAAGACCAGGGCATA 1320
Db       1328 GGCACAGACTGCACAAGCATAGGGACAATAGAGCAGAAGTCCGGAGAAGACCAGGGCATA 1387

Qy       1321 AAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCAAGAAAAAACTCAAGATATTCCAG 1380
Db       1388 AAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCAAGAAAAAACTCAAGATATTCCAG 1447

Qy       1381 CCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCA 1440
Db       1448 CCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCA 1507

Qy       1441 CAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATG 1500
Db       1508 CAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATG 1567

Qy       1501 AGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAACCCAAGGAAAAGGTCAAGACGAAG 1560
Db       1568 AGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAACCCAAGGAAAAGGTCAAGACGAAG 1627

Qy       1561 CCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTGGGGATTAAAATCTCTTCTGTG 1620
Db       1628 CCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTGGGGATTAAAATCTCTTCTGTG 1687

Qy       1621 CACAAGAGACTAGCGTCAGAGAAAAGGGAAAACCCAGTGAAGAAAGTGATGCTGGCTTCC 1680
Db       1688 CACAAGAGACTAGCGTCAGAGAAAAGGGAAAACCCAGTGAAGAAAGTGATGCTGGCTTCC 1747

Qy       1681 AGGAGTGAGACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGCACACCATCCTGGGCAAGT 1740
Db       1748 AGGAGTGAGACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGCACACCATCCTGGGCAAGT 1807

Qy       1741 GACCACAACTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAGCCCACTGCACTCTCGCCC 1800
Db       1808 GACCACAACTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAGCCCACTGCACTCTCGCCC 1867

Qy       1801 ACCCTATTGAGTAAATGTACGTATCACCCAAAGGCTGGCTTCCCAGGCCCCTCCCATCAT 1860
Db       1868 ACCCTATTGAGTAAATGTACGTATCACCCAAAGGCTGGCTTCCCAGGCCCCTCCCATCAT 1927

Qy       1861 CTGGGTGGCTGCCTGGGGCTGTCTAGGACCAGAGTCCTGGGTGTTCTGGTGCTGATAGTA 1920
Db       1928 CTGGGTGGCTGCCTGGGGCTGTCTAGGACCAGAGTCCTGGGTGTTCTGGTGCTGATAGTA 1987

Qy       1921 GCCAGCAGCTCACTGCCAGCCAGAAGCAGATACCAAGATGCCTCTGGACCCCAGGTGTTC 1980
Db       1988 GCCAGCAGCTCACTGCCAGCCAGAAGCAGATACCAAGATGCCTCTGGACCCCAGGTGTTC 2047

Qy       1981 CTGCCTAGCCTGTGGAGCCTCTCTGGGTGGTTCCTAAAGAGCTGTGTAGGCCTCATGTTG 2040
Db       2048 CTGCCTAGCCTGTGGAGCCTCTCTGGGTGGTTCCTAAAGAGCTGTGTAGGCCTCATGTTG 2107

Qy       2041 GAGGCAATTTCTTATTTCAGTTTTAGGCCTTGGTAAGTATTTGAACTGCTCTCAACAAGA 2100
Db       2108 GAGGCAATTTCTTATTTCAGTTTTAGGCCTTGGTAAGTATTTGAACTGCTCTCAACAAGA 2167

Qy       2101 TGGGACAACATCAGTGCACTGTTCAGACTCAGTTCAGACTTGAGTTCCTCACAGGACAGC 2160
Db       2168 TGGGACAACATCAGTGCACTGTTCAGACTCAGTTCAGACTTGAGTTCCTCACAGGACAGC 2227

Qy       2161 CCAGCACAAACCATGGAGTTTTCCCACGTTACCTATTTTTCTTCAACATGGTCACCACTG 2220
Db       2228 CCAGCACAAACCAT-GAGTTTTCCCACGTTACCTATTTTCTTC--ACATGGTCACCACTG 2284

Qy       2221 CTCTTCTATAGAAAAGTGATTTTTTTCATGAGTTAGGGCATTTTTGAAAGGACAGGTAGA 2280
Db       2285 CTCTTCTATAGAAAAGTGATTTTTTTCATGAGTTAGGGCATTTTTGAAAGGACAGGTAGA 2344

Qy       2281 CAATTTGTGTCTATTTCACATACTACACACCTACATAAACAGGTTTGAATTTTGAAGGTC 2340
Db       2345 CAATTTGTGTCTATTTCACATACTACACACCTACATAAACAGGTTTGAATTTTGAAGGTC 2404

Qy       2341 ATCTGCTGCTCAGATCCCATGGTGAGTAATCGTGCTGACCATACACCTACAGCAGTATTC 2400
Db       2405 ATCTGCTGCTCAGATCCCATGGTGAGTAATCGTGCTGACCATACACCTACAGCAGTATTC 2464

Qy       2401 TCATCAAAGTGAGAATTGTGGGTGGTAGCTTCCAGACTCGTAGATGCCCCAGATTTCAAA 2460
Db       2465 TCATCAAAGTGAGAATTGTGGGTGGTAGCTTCCAGACTCGTAGATGCCCCAGATTTCAAA 2524

Qy       2461 TCAGTGTTTATAGTTTGAGTAAACTTTTGAAACCCAAGATTAGATCACTGAGTTTGGAGC 2520
Db       2525 TCAGTGTTTATAGTTTGAGTAAACTTTTGAAACCCAAGATTAGATCACTGAGTTTGGAGC 2584

Qy       2521 CATTCTTGTTTCACTGCTGTACAAGGTACACCTTTCTTACTTTTTGTACAGTGGTTTCAA 2580
Db       2585 CATTCTTGTTTCACTGCTGTACAAGGTACACCTTTCTTACTTTTTGTACAGTGGTTTCAA 2644

Qy       2581 GTGAATTTTCATCTGTGTTCTCATGATAGTGTTTTGTTCCATCTCTGTCCACCCTCAGCC 2640
Db       2645 GTGAATTTTCATCTGTGTTCTCATGATAGTGTTTTGTTCCATCTCTGTCCACCCTCAGCC 2704

Qy       2641 CCAGATAAGGGGGTCTCATGGCCTAAGCTGGCCTTAAATTTTCTGTAAAGTTGGGGGTCC 2700
Db       2705 CCAGATAAGGGGGTCTCATGGCCTAGGCTGGCCTTAAATTTTCTGTAAAGTTGGGGGTCC 2764

Qy       2701 TCCTGCCCCCACCTGTAAAGAGTGCTGAAACTTACAGGCACACACCTGATGTATGCTGTG 2760
Db       2765 TCCTGCCCCCACCTGTAAAGAGTGCTGAAACTTACAGGCACACACCTGATGTATGTTGTG 2824

Qy       2761 GGGTGTGAACTCAGAGTGTTGGACCAGCAGTCTACCAGCTGAGCTGCAGTTCTAGCCATG 2820
Db       2825 GGGTGTGAACTCAGAGTGTTGGACCAGCAGTCTACCAGCTGAGCTGCAGTTCTAGCCATG 2884

Qy       2821 GTAACTGAAACTGCTGATCTTCTTGTCTTTGTTTCCTTTGTGCTGGG 2867
Db       2885 GTAACTGAAACTGCTGATCTTCTTGTCTTTGTTTCCTTTGTGCTGGG 2931
```

RESULT 2
BC060249
LOCUS BC060249 4651 bp mRNA linear HTC 19-NOV-2003
DEFINITION Mus musculus death inducer-obliterator 1, mRNA (cDNA clone IMAGE:6826043), containing frame-shift errors.

ACCESSION BC060249
VERSION BC060249.1 GI:37805411
KEYWORDS HTC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 4651)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.P., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE 22388257
PUBMED 12477932
REFERENCE 2 (bases 1 to 4651)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (20-OCT-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: M. Bento Soares, University of Iowa
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 126 Row: k Column: 20
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein
This clone has the following problem: frame shifted.

FEATURES Location/Qualifiers
source 1. .4651
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6826043"
/tissue_type="Brain, mouse, 13.5,14.5,16.5,17.5 dpc"
/clone_lib="NIH_BMAP_FY0"
/lab_host="DH10B"
/note="Vector: pYX-ASC"

ORIGIN

Query Match 96.8%; Score 2775.8; DB 11; Length 4651;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2814; Conservative 0; Mismatches 7; Indels 4; Gaps 3;

```
Qy      44 GGAGCTTGTTTAAGAGGCAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTT 103
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      90 GGAGCTTGTTTAAGAGGCAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTT 149

Qy     104 CTCATTGATGAACTGTGACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGG 163
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     150 CTCATTGATGAACTGTGACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGG 209

Qy     164 CCCCAGGTCTATAATTTTATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTG 223
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     210 CCCCAGGTCTATAATTTTATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTG 269

Qy     224 GGACAGGTATGGATGATAAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAAC 283
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     270 GGACAGGTATGGATGATAAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAAC 329

Qy     284 CCACCAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTG 343
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     330 CCACCAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTG 389

Qy     344 AGGGTGCAGGAGACACGGAGGCGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCT 403
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     390 AGGGTGCAGGAGACACGGAGGTGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCT 449

Qy     404 CCCTGCGCCGCAGTGGACGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCA 463
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     450 CCCTGCGCCGCAGTGGACGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCA 509

Qy     464 CGGTTCGGCGCCGAGGGAAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCA 523
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     510 CGGTTCGGCGCCGAGGGAAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCA 569

Qy     524 CATCTTCCACAGTCACTGATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTG 583
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     570 CATCTTCCACAGTCACTGATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTG 629

Qy     584 AGATCAGAAGTGGCCCTGTATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAAA 643
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     630 AGATCAGAAGTGGCCCTGTATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAAA 689

Qy     644 AGGCAAAAGGAGGTGAAGAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACGT 703
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     690 AGGCAAAAGGAGGTGAAGAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACGT 749

Qy     704 TGAAGGAACTTCAGAACCGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCCC 763
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     750 TGAAGGAACTTCAGAACCGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCCC 809

Qy     764 TGAGAGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTG 823
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     810 TGAGAGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTG 869

Qy     824 GGAGTGTCCAAATAGGCAGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGG 883
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     870 GGAGTGTCCAAATAGGCAGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGG 929

Qy     884 CTAGTCAGGGACCAGTGTCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGA 943
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     930 CTAGTCAGGGACCAGTGTCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGA 989

Qy     944 AGGCGACTCAGGGAAATACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAGT 1003
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     990 AGGCGACTCAGGGAAATACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAGT 1049

Qy    1004 GTGAGGTTTACGACCCCAATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGT 1063
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      1050 GTGAGGTTTACGACCCCAATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGT 1109

Qy      1064 TTATGATCTGCTGTGATCGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCTG 1123

Db      1110 TTATGATCTGCTGTGATCGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCTG 1169

Qy      1124 AGGCCCGAGGGCGGCTCCTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTGCACCA 1183

Db      1170 AGGCCCGAGGGCGGCTCCTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTGCACCA 1229

Qy      1184 TTTTGCAAGTGCAGGATGAGACAAACGGTAGCGCCACCAATGAGCAGGACTCTGGGTGCA 1243

Db      1230 TTTTGCAAGTGCAGGATGAGACAAACGGTAGCGCCACCGATGAGCAGGACTCTGGGTGCA 1289

Qy      1244 GATCTGTGGGTGCTGATGGCACAGACTGCACAAGCATAGGGACAGTAGAGCAGAAGTCCG 1303

Db      1290 GATCTGTGGGTGCTGATGGCACAGACTGCACAAGCATAGGGACAGTAGAGCAGAAGTCCG 1349

Qy      1304 GAGAAGACCAGGGCATAAAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCAAGAAAA 1363

Db      1350 GAGAAGACCAGGGCATAAAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCAAGAAAA 1409

Qy      1364 AACTCAAGATATTCCAGCCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTGGCCCTG 1423

Db      1410 AACTCAAGATATTCCAGCCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTGGCCCTG 1469

Qy      1424 GGTGTTCCAGTGTAGCACAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTCTCAAAC 1483

Db      1470 GGTGTTCCAGTGTAGCACAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTCTCAAAC 1529

Qy      1484 ACGCAGCAGCTACCATGAGATTTCTAAGTTCAGGTAAAGAAC-AAAAAACAAAACCCAAG 1542

Db      1530 ACGCAGCAGCTACCATGAGATTTCTAAGTTCAGGTAAAGAACAAAAAAACAAAACCCAAG 1589

Qy      1543 GAAAAGGTCAAGACGAAGCCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTGGGG 1602

Db      1590 GAAAAGGTCAAGACGAAGCCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTGGGG 1649

Qy      1603 ATTAAAATCTCTTCTGTGCACAAGAGACTAGCGTCAGAGAAAAGGGAAAACCCAGTGAAG 1662

Db      1650 ATTAAAATCTCTTCTGTGCACAAGAGACTAGCGTCAGAGAAAAGGGAAAACCCAGTGAAG 1709

Qy      1663 AAAGTGATGCTGGCTTCCAGGAGTGAGACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGC 1722

Db      1710 AAAGTGATGCTGGCTTCCAGGAGTGAGACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGC 1769

Qy      1723 ACACCATCCTGGGCAAGTGACCACAACTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAG 1782

Db      1770 ACACCATCCTGGGCAAGTGACCACAACTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAG 1829

Qy      1783 CCCACTGCACTCTCGCCCACCCTATTGAGTAAATGTACGTATCACCCAAAGGCTGGCTTC 1842

Db      1830 CCCACTGCACTCTCGCCCACCCTATTGAGTAAATGTACGTATCACCCAAAGGCTGGCTTC 1889

Qy      1843 CCAGGCCCCTCCCATCATCTGGGTGGCTGCCTGGGGCTGTCTAGGACCAGAGTCCTGGGT 1902

Db      1890 CCAGGCCCCTCCCATCATCTGGGTGGCTGCCTGGGGCTGTCTAGGACCAGAGTCCTGGGT 1949

Qy      1903 GTTCTGGTGCTGATAGTAGCCAGCAGCTCACTGCCAGCCAGAAGCAGATACCAAGATGCC 1962

Db      1950 GTTCTGGTGCTGATAGTAGCCAGCAGCTCACTGCCAGCCAGAAGCAGATACCAAGATGCC 2009

Qy      1963 TCTGGACCCCAGGTGTTCCTGCCTAGCCTGTGGAGCCTCTCTGGGTGGTTCCTAAAGAGC 2022

Db      2010 TCTGGACCCCAGGTGTTCCTGCCTAGCCTGTGGAGCCTCTCTGGGTGGTTCCTAAAGAGC 2069

Qy      2023 TGTGTAGGCCTCATGTTGGAGGCAATTTCTTATTTCAGTTTTAGGCCTTGGTAAGTATTT 2082

Db      2070 TGTGTAGGCCTCATGTTGGAGGCAATTTCTTATTTCAGTTTTAGGCCTTGGTAAGTATTT 2129

Qy      2083 GAACTGCTCTCAACAAGATGGGACAACATCAGTGCACTGTTCAGACTCAGTTCAGACTTG 2142

Db      2130 GAACTGCTCTCAACAAGATGGGACAACATCAGTGCACTGTTCAGACTCAGTTCAGACTTG 2189

Qy      2143 AGTTCCTCACAGGACAGCCCAGCACAAACCATGGAGTTTTCCCACGTTACCTATTTTTCT 2202

Db      2190 AGTTCCTCACAGGACAGCCCAGCACAAACCAT-GAGTTTTCCCACGTTACCTATTTT--C 2246

Qy      2203 TCAACATGGTCACCACTGCTCTTCTATAGAAAAGTGATTTTTTTCATGAGTTAGGGCATT 2262

Db      2247 TTCACATGGTCACCACTGCTCTTCTATAGAAAAGTGATTTTTTTCATGAGTTAGGGCATT 2306

Qy      2263 TTTGAAAGGACAGGTAGACAATTTGTGTCTATTTCACATACTACACACCTACATAAACAG 2322

Db      2307 TTTGAAAGGACAGGTAGACAATTTGTGTCTATTTCACATACTACACACCTACATAAACAG 2366

Qy      2323 GTTTGAATTTTGAAGGTCATCTGCTGCTCAGATCCCATGGTGAGTAATCGTGCTGACCAT 2382

Db      2367 GTTTGAATTTTGAAGGTCATCTGCTGCTCAGATCCCATGGTGAGTAATCGTGCTGACCAT 2426

Qy      2383 ACACCTACAGCAGTATTCTCATCAAAGTGAGAATTGTGGGTGGTAGCTTCCAGACTCGTA 2442

Db      2427 ACACCTACAGCAGTATTCTCATCAAAGTGAGAATTGTGGGTGGTAGCTTCCAGACTCGTA 2486

Qy      2443 GATGCCCCAGATTTCAAATCAGTGTTTATAGTTTGAGTAAACTTTTGAAACCCAAGATTA 2502

Db      2487 GATGCCCCAGATTTCAAATCAGTGTTTATAGTTTGAGTAAACTTTTGAAACCCAAGATTA 2546

Qy      2503 GATCACTGAGTTTGGAGCCATTCTTGTTTCACTGCTGTACAAGGTACACCTTTCTTACTT 2562

Db      2547 GATCACTGAGTTTGGAGCCATTCTTGTTTCACTGCTGTACAAGGTACACCTTTCTTACTT 2606

Qy      2563 TTTGTACAGTGGTTTCAAGTGAATTTTCATCTGTGTTCTCATGATAGTGTTTTGTTCCAT 2622

Db      2607 TTTGTACAGTGGTTTCAAGTGAATTTTCATCTGTGTTCTCATGATAGTGTTTTGTTCCAT 2666

Qy      2623 CTCTGTCCACCCTCAGCCCCAGATAAGGGGGTCTCATGGCCTAAGCTGGCCTTAAATTTT 2682

Db      2667 CTCTGTCCACCCTCAGCCCCAGATAAGGGGGTCTCATGGCCTAGGCTGGCCTTAAATTTT 2726

Qy      2683 CTGTAAAGTTGGGGGTCCTCCTGCCCCCACCTGTAAAGAGTGCTGAAACTTACAGGCACA 2742

Db      2727 CTGTAAAGTTGGGGGTCCTCCTGCCCCCACCTGTAAAGAGTGCTGAAACTTACAGGCACA 2786

Qy      2743 CACCTGATGTATGCTGTGGGGTGTGAACTCAGAGTGTTGGACCAGCAGTCTACCAGCTGA 2802

Db      2787 CACCTGATGTATGTTGTGGGGTGTGAACTCAGAGTGTTGGACCAGCAGTCTACCAGCTGA 2846

Qy      2803 GCTGCAGTTCTAGCCATGGTAACTGAAACTGCTGATCTTCTTGTCTTTGTTTCCTTTGTG 2862

Db      2847 GCTGCAGTTCTAGCCATGGTAACTGAAACTGCTGATCTTCTTGTCTTTGTTTCCTTTGTG 2906

Qy      2863 CTGGG 2867

Db      2907 CTGGG 2911
```

RESULT 3
AK014422

LOCUS	AK014422 1614 bp mRNA linear HTC 20-SEP-2003
DEFINITION	Mus musculus 18 days pregnant adult female placenta and extra embryonic tissue cDNA, RIKEN full-length enriched library, clone:3830408B01 product:death inducer-obliterator-1, full insert sequence.
ACCESSION	AK014422
VERSION	AK014422.1 GI:12852263
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1
AUTHORS	Carninci,P. and Hayashizaki,Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
MEDLINE	99279253

```
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 1614)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
            Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,
            Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
            Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
            Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
            Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
            Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
            Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K.,
            Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 10.0 and subtraction to
            Rot = 50.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
            with BamHI and XhoI. Vector: a modified pBluescript KS(+) after
            bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3'
            end: BamHI. Host: DH10B.
FEATURES            Location/Qualifiers
     source         1. .1614
                    /organism="Mus musculus"
                    /mol_type="mRNA"
                    /strain="C57BL/6J"
                    /db_xref="FANTOM_DB:3830408B01"
                    /db_xref="MGI:1910351"
                    /db_xref="taxon:10090"
                    /clone="3830408B01"
                    /sex="female"
                    /tissue_type="placenta and extra embryonic tissue"
                    /clone_lib="RIKEN full-length enriched mouse cDNA library"
                    /dev_stage="18 days pregnant adult"
     misc_feature   331. .>1614
                    /note="death inducer-obliterator-1 (MGD|MGI:1344352,
                    GB|AJ238332, evidence: BLASTN, 99%, match=2599)
                    putative"
ORIGIN

  Query Match             51.7%;  Score 1482.4;  DB 11;  Length 1614;
  Best Local Similarity   99.4%;  Pred. No. 0;
  Matches 1509;  Conservative    0;  Mismatches    6;  Indels    3;  Gaps    2;

Qy          1 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        100 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 159

Qy         61 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        160 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 219

Qy        121 ACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        220 ACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTT 279

Qy        181 TATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        280 TATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGAT 339

Qy        241 AAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        340 AAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTC 399

Qy        301 AGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 AGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACG 459

Qy        361 GAGGCGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGA 420
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 GAGGTGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGA 519

Qy        421 CGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 CGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGG 579

Qy        481 AAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 AAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACT 639

Qy        541 GATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCT 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        640 GATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGT-GCCCT 698

Qy        601 GTATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAA 660
              |||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        699 GTATCTGACTCCTTAGG--AAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAA 756

Qy        661 GAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAAC 720
              |||||||||||||||||||||||||||||||||||||| |||||||| ||||||||||||
Db        757 GAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCTTTACGTTGAGGGAACTTCAGAAC 816

Qy        721 CGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAAT 780
              |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db        817 CGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGCTCCCTGAGAGGCAGTCAGAAT 876
```

```
Qy        781 CGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        877 CGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGC 936

Qy        841 AGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        937 AGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTG 996

Qy        901 TCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        997 TCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAAT 1056

Qy        961 ACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1057 ACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCC 1116

Qy       1021 AATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1117 AATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGAT 1176

Qy       1081 CGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1177 CGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTC 1236

Qy       1141 CTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGAT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1237 CTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGAT 1296

Qy       1201 GAGACAAACGGTAGCGCCACCAATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGAT 1260
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       1297 GAGACAAACGGTAGCGCCACCGATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGAT 1356

Qy       1261 GGCACAGACTGCACAAGCATAGGGACAGTAGAGCAGAAGTCCGGAGAAGACCAGGGCATA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1357 GGCACAGACTGCACAAGCATAGGGACAGTAGAGCAGAAGTCCGGAGAAGACCAGGGCATA 1416

Qy       1321 AAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCAAGAAAAAACTCAAGATATTCCAG 1380
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       1417 AAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCAAAAAAAAACTCAAGATATTCCAG 1476

Qy       1381 CCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1477 CCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCA 1536

Qy       1441 CAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1537 CAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATG 1596

Qy       1501 AGATTTCTAAGTTCAGGT 1518
              ||||||||||||||||||
Db       1597 AGATTTCTAAGTTCAGGT 1614
```

RESULT 4
BC029110
LOCUS BC029110 1640 bp mRNA linear HTC 19-NOV-2003
DEFINITION Mus musculus death inducer-obliterator 1, mRNA (cDNA clone IMAGE:3980050), containing frame-shift errors.
ACCESSION BC029110
VERSION BC029110.1 GI:22137689
KEYWORDS HTC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1640)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE 22388257
PUBMED 12477932
REFERENCE 2 (bases 1 to 1640)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (01-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 23 Row: k Column: 3
This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein
This clone has the following problem: frame shifted.
FEATURES Location/Qualifiers
source 1. .1640
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3980050"
/tissue_type="Mammary tumor. Metallothionien-TGP alpha model. 10 month old virgin mouse. Taken by biopsy."
/clone_lib="NCI_CGAP_Mam1"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
ORIGIN

Query Match 48.8%; Score 1398.8; DB 11; Length 1640;
Best Local Similarity 93.0%; Pred. No. 0;
Matches 1522; Conservative 0; Mismatches 2; Indels 112; Gaps 1;

```
Qy          13 CTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGGCAGTCCCCACTG 72
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 CTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGGCAGTCCCCACTG 60

Qy          73 TCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTGACTGACAACAGT 132
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 TCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTGACTGACAACAGT 120

Qy         133 GGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTTTATGTAGGATTT 192
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       121 GGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTTTATGTAGGATTT 180
Qy       193 CCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGAC------------------------- 227
             ||||||||||||||||||||||||||||||||||
Db       181 CCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGGTTTTTGTTTTTATGCCAGTCT 240
Qy       228 ------------------------------------------------------------ 227
Db       241 CCCCCTAAGATGGAACATGCAGTAAACTTGTCCATTACTAGCATTTAAATGCAGTTAAAG 300
Qy       228 ---------------------------AGGTATGGATGATAAAGGGCACCTGAGCAATGA 260
                                         ||||||||||||||||||||||||||||||||||
Db       301 CCTGGACAGCCAGGAAAAGAATTCTCAAGGTATGGATGATAAAGGGCACCTGAGCAATGA 360
Qy       261 GGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCG 320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 GGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCG 420
Qy       321 AAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACGGAGGCGGACCCCAGTGAGCA 380
             |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       421 AAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACGGAGGTGGACCCCAGTGAGCA 480
Qy       381 GCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGACGGCAACCAAAACGTACTGA 440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 GCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGACGGCAACCAAAACGTACTGA 540
Qy       441 GAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGGAAAAAGAATGTGCCGGTGTC 500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 GAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGGAAAAAGAATGTGCCGGTGTC 600
Qy       501 CCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACTGATGTGGAGACAGCTTCCGA 560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 CCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACTGATGTGGAGACAGCTTCCGA 660
Qy       561 GGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCTGTATCTGACTCCTTAGGGAA 620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 GGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCTGTATCTGACTCCTTAGGGAA 720
Qy       621 AGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAAGAGGAAGAAGACACCTCTGA 680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 AGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAAGAGGAAGAAGACACCTCTGA 780
Qy       681 CAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAACCGCCTTCGGAGAAAGCGAGA 740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 CAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAACCGCCTTCGGAGAAAGCGAGA 840
Qy       741 GCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAG 800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 GCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAG 900
Qy       801 AGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGCAGTGCCGAGCAGGACAGACC 860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 AGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGCAGTGCCGAGCAGGACAGACC 960
Qy       861 TCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTGTCCCAGTCAGAGACAGATGA 920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 TCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTGTCCCAGTCAGAGACAGATGA 1020
Qy       921 CATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAATACAGAGGAAAACCCCAGGGA 980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 CATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAATACAGAGGAAAACCCCAGGGA 1080
Qy       981 AGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCCAATGCCCTGTACTGCATCTG 1040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 AGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCCAATGCCCTGTACTGCATCTG 1140
Qy      1041 CCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGATCGGTGTGAGGAGTGGTTCCA 1100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 CCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGATCGGTGTGAGGAGTGGTTCCA 1200
Qy      1101 TGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAACGGGGAAGA 1160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 TGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAACGGGGAAGA 1260

Qy      1161 CTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGATGAGACAAACGGTAGCGCCAC 1220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 CTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGATGAGACAAACGGTAGCGCCAC 1320
Qy      1221 CAATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGATGGCACAGACTGCACAAGCAT 1280
             | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 CGATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGATGGCACAGACTGCACAAGCAT 1380
Qy      1281 AGGGACAGTAGAGCAGAAGTCCGGAGAAGACCAGGGCATAAAGGGTAGGATTGAGAAGGC 1340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 AGGGACAGTAGAGCAGAAGTCCGGAGAAGACCAGGGCATAAAGGGTAGGATTGAGAAGGC 1440
Qy      1341 AGCAAACCCCAGCGGCAAGAAAAAACTCAAGATATTCCAGCCTGTCGTAGAGGCTCCTGG 1400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 AGCAAACCCCAGCGGCAAGAAAAAACTCAAGATATTCCAGCCTGTCGTAGAGGCTCCTGG 1500
Qy      1401 TGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCACAGCCTGACTCTGTGTATTG 1460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1501 TGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCACAGCCTGACTCTGTGTATTG 1560
Qy      1461 CAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATGAGATTTCTAAGTTCAGGTAA 1520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1561 CAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATGAGATTTCTAAGTTCAGGTAA 1620
Qy      1521 AGAACAAAAAACAAAA 1536
             ||||||||||||||||
Db      1621 AGAACAAAAAACAAAA 1636
```

RESULT 5
BC014615

LOCUS	BC014615 2752 bp mRNA linear HTC 17-DEC-2003
DEFINITION	Homo sapiens cDNA clone IMAGE:3532298, containing frame-shift errors.
ACCESSION	BC014615
VERSION	BC014615.2 GI:33869935
KEYWORDS	HTC.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 2752)
AUTHORS	Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED	12477932
REFERENCE	2 (bases 1 to 2752)
AUTHORS	Strausberg,R.
TITLE	Direct Submission
JOURNAL	Submitted (24-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov

```
COMMENT     On Aug 19, 2003 this sequence version replaced gi:15779097.
            Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 11 Row: j Column: 11
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 18375616
            This clone has the following problem: frame shifted.
FEATURES             Location/Qualifiers
     source          1. .2752
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGB:3532298"
                     /tissue_type="Muscle, rhabdomyosarcoma"
                     /clone_lib="NIH_MGC_17"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
ORIGIN

  Query Match           35.3%;  Score 1010.8;  DB 11;  Length 2752;
  Best Local Similarity  71.8%;  Pred. No. 3.2e-236;
  Matches 1481;  Conservative    0;  Mismatches  532;  Indels   51;  Gaps   10;

Qy         16 GAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGGCAGTCCCCACTGTCC 75
Db          1 GAAATGGCTGCGAGACCCTAGAGGCCTGCGGAGCTTACTCCACGGGAACAGCCTCTAGAT 60

Qy         76 CTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTGACTGACAACAGTGGG 135
Db         61 AATCTGAGTTGTTGAAAATACGAAGCCTGTTACTCGTGAACAGTGGCTGACAACAGTGTT 120

Qy        136 GT--GAGGCTTGGCCGTCTGCTTG----------ACCTGGCCCCAGGTCTATAATTTT 181
Db        121 GTTGTGAGCCTGGCTGTCTGCTTGGACCCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGT 180

Qy        182 ATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGATA 241
Db        181 ATTTAGGATTT-CAGGGAAAAGTGTCCAAGCTTTCAGTGTTGGAGCAGGTATGGACGACA 239

Qy        242 AAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTCA 301
Db        240 AAGGCGACCCGAGCAATGAGGAGGCACCTAAGGCCATCAAACCCACCAGCAAAGAGTTCA 299

Qy        302 GGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACGG 361
Db        300 GGAAAACATGGGGTTTTCGAAGGACCACTATCGCCAAGCGAGAGGGCGCAGGGGACGCGG 359

Qy        362 AGGCGGACCCCAGTGAGCAGCAACC------ACAGCAGCATAACCTCTCCCTGCGCCGCA 415
Db        360 AGGCTGACCCACTGGAGCCGCCACCCCCACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCA 419

Qy        416 GTGGACGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCC 475
Db        420 GTGGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCC 479

Qy        476 GAGGGAAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAG 535
Db        480 GCGGCAGGAGGAGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCG 539

Qy        536 TCACTGATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTG 595
Db        540 CCACAGACGCCGAGACAGCCTCCGAGGGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCG 599

Qy        596 GCCCTGTATCTGACTCCTTA---GGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAG 652
Db        600 GCCCCCAGTCTGCTTCCACAGCTGTGAAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAG 659

Qy        653 GAGGTGAAGAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGAAC 712
Db        660 GAGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGGCCTGACCTTGAAAGAGC 719

Qy        713 TTCAGAACCGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCA 772
Db        720 TTCAGAATCGCCTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGA 779

Qy        773 GTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCC 832
Db        780 TCCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCG 839

Qy        833 AAATAGGCAGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGG 892
Db        840 AGGCCAGTGACACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGG 899

Qy        893 GACCAGTGTCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTC 952
Db        900 GGGTTGTGTCCCAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTC 959

Qy        953 AGGGAAATACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTT 1012
Db        960 AGGACATCAAAGATGAGGAGCCTGGAGACTTGGGCCGACCGAAGCCTGAATGTGAGGGTT 1019

Qy       1013 ACGACCCCAATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCT 1072
Db       1020 ACGACCCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTT 1079

Qy       1073 GCTGTGATCGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAG 1132
Db       1080 GCTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAG 1139

Qy       1133 GGCGGCTCCTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTGCACCATTTTGCAAG 1192
Db       1140 GGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACTGCACCATTCTGCAAG 1199

Qy       1193 TGCAGGATGAGACAAACGGTAGCGCCACCAATGAGCAGGACTCTGGGTGCAGATCTGTGG 1252
Db       1200 TGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAG 1259

Qy       1253 GTGCTGATGGCACAGACTGCACAAGCATAGGGACAGTAGAGCAGAAGTCCGGAGAAGACC 1312
Db       1260 ATGCTGATGGCACCGATTGTACAAGTATAGGAACAATAGAGCAGAAGTCTAGCGAAGACC 1319

Qy       1313 AGGGCATAAAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCAAGAAAAAACTCAAGA 1372
Db       1320 AAGGGATAAAGGGTAGAATTGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGA 1379

Qy       1373 TATTCCAGCCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCCA 1432
Db       1380 TCTTCCAGCCTGTGATAGAGGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTC 1439

Qy       1433 GTGTAGCACAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTCTCAAACACGCAGCAG 1492
Db       1440 ACGTGGCGCAGCCCGACTCGGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAG 1499

Qy       1493 CTACCATGAGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAACCCAAGGAAAAGGTCA 1552
Db       1500 CGACAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGA 1559

Qy       1553 AGACGAAGCCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTGGGGATTAAAATCT 1612
Db       1560 AGATGAAGCCAGAGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCT 1619

Qy       1613 CTTCTGTGCACAAGAGACTAGCGTCAGAGAAAAGGGAAAACC----CAGTGAAGAAAGTG 1668
Db       1620 CTTCTGTGCACAAGAGACCAGCTCCAGAAAAAAAAAGAGACCACAGTGAAGAAGGCAGTG 1679
```

```
Qy       1669 ATGCTGGCTTCCAGGAGTGAGACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGCACACCA 1728
Db       1680 GTGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCG 1739

Qy       1729 TCCTGGGCAAGTGACCACAACTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAGCCCACT 1788
Db       1740 TCGTGGGCGAGCGATCACAATTACAATGCAGTAAAGCCAGAAAAGACTG---------CT 1790

Qy       1789 GCACTCTCGCCCACCCTATTGAGTAAATGTACGTATCACCCAAAGGCTGGCTTCCCAGGC 1848
Db       1791 GCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACCTAGGGGTTGGCCTCCTGGAC 1850

Qy       1849 CCCTCCCATCATCT-----GGGTGGCTGCCTGGGGCTGTCTAGGACCAG-----AGTCCT 1898
Db       1851 CCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCCTGTCCAGGACTGGGAGTTGCAGCT 1910

Qy       1899 GGGTGTTCTGGTGCTGATAGTAGCCAGCAGC----TCACTGCCAGCCAGAAGCAGATACC 1954
Db       1911 TTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCAGCGGGAAGCAGAGCCCATGTCC 1970

Qy       1955 AAGATGCCTCTGGACCCCAGGTGTTCCTGCCTAGCCTGTGGAGCCTCTCTGGGTGGTTCC 2014
Db       1971 AGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTCTGTCAGGACTTCCTGTCACTGTTT 2030

Qy       2015 TAAAGAGCTGTGTAGGCCTCATGT 2038
Db       2031 TCCAAAGCTGTAAACCTCACTGGT 2054
```

RESULT 6
AK044919
LOCUS AK044919 3385 bp mRNA linear HTC 20-SEP-2003
DEFINITION Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length enriched library, clone:B130011M08 product:death inducer-obliterator-1, full insert sequence.
ACCESSION AK044919
VERSION AK044919.1 GI:26336914
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 3385)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by Dr. Tomohiro Kono (Department of Animal Science, Tokyo University of Agriculture, 1737 Hunako Atsugi City, Kanagawa Prefecture, Japan) whose assistance we gratefully acknowledge.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/.
FEATURES Location/Qualifiers
source 1. .3385
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:B130011M08"
/db_xref="MGI:2410273"
/db_xref="taxon:10090"
/clone="B130011M08"
/tissue_type="parthenogenote"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="9.5 days embryo"
CDS 2443. .3276
/note="unnamed protein product; death inducer-obliterator-1 (MGD|MGI:1344352, GB|AJ238332, evidence: BLASTN, 99%, match=2599) putative"
/codon_start=1
/protein_id="BAC32141.1"
/db_xref="GI:26336915"
/translation="MDDKGHLSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTEVDPSEQQPQQHNLSLRRSGRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDSLGKEHPASSEKAKGGEEEEDTSDSDSDGLTLKELQNRLRRKREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSAEQDRPLCKQEPEASQGPVSQSETDDIENQLEGKATQGNTEENPREAGKPKPECEVYDPNALYCICRQPHNNR"
ORIGIN

```
Query Match 29.1%; Score 835.6; DB 11; Length 3385;
Best Local Similarity 98.4%; Pred. No. 2.6e-193;
Matches 844; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy       211 GCTTTCAGTTTTGGGACAGGTATGGATGATAAAGGGCACCTGAGCAATGAGGAAGCACCC 270
```

```
Db       2422 GATTGCCTTTTGTCATTAGGTATGGATGATAAAGGGCACCTGAGCAATGAGGAAGCACCC 2481

Qy        271 AAGGCTATCAAACCCACCAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCGAAGAACCACG 330

Db       2482 AAGGCTATCAAACCCACCAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCGAAGAACCACG 2541

Qy        331 ATTGCCAAACGTGAGGGTGCAGGAGACACGGAGGCGGACCCCAGTGAGCAGCAACCACAG 390

Db       2542 ATTGCCAAACGTGAGGGTGCAGGAGACACGGAGGTGGACCCCAGTGAGCAGCAACCACAG 2601

Qy        391 CAGCATAACCTCTCCCTGCGCCGCAGTGGACGGCAACCAAAACGTACTGAGAGGGTAGAA 450

Db       2602 CAGCATAACCTCTCCCTGCGCCGCAGTGGACGGCAACCAAAACGTACTGAGAGGGTAGAA 2661

Qy        451 GAGTTTCTTACCACGGTTCGGCGCCGAGGGAAAAAGAATGTGCCGGTGTCCCTGGAGGAT 510

Db       2662 GAGTTTCTTACCACGGTTCGGCGCCGAGGGAAAAAGAATGTGCCGGTGTCCCTGGAGGAT 2721

Qy        511 TCCAGTGAGCCCACATCTTCCACAGTCACTGATGTGGAGACAGCTTCCGAGGGGAGCGTT 570

Db       2722 TCCAGTGAGCCCACATCTTCCACAGTCACTGATGTGGAGACAGCTTCCGAGGGGAGCGTT 2781

Qy        571 GAAAGCAGTTCTGAGATCAGAAGTGGCCCTGTATCTGACTCCTTAGGGAAAGAACATCCT 630

Db       2782 GAAAGCAGTTCTGAGATCAGAAGTGGCCCTGTATCTGACTCCTTAGGGAAAGAACATCCT 2841

Qy        631 GCCTCTTCTGAAAAGGCAAAAGGAGGTGAAGAGGAAGAAGACACCTCTGACAGTGACAGT 690

Db       2842 GCCTCTTCTGAAAAGGCAAAAGGAGGTGAAGAGGAAGAAGACACCTCTGACAGTGACAGT 2901

Qy        691 GATGGCCTTACGTTGAAGGAACTTCAGAACCGCCTTCGGAGAAAGCGAGAGCAAGAACCT 750

Db       2902 GATGGCCTTACGTTGAAGGAACTTCAGAACCGCCTTCGGAGAAAGCGAGAGCAAGAACCT 2961

Qy        751 GTGGAGAGGTCCCTGAGAGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAAGAT 810

Db       2962 GTGGAGAGGTCCCTGAGAGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAAGAT 3021

Qy        811 TCTGCCGAAACTGGGAGTGTCCAAATAGGCAGTGCCGAGCAGGACAGACCTCTCTGTAAG 870

Db       3022 TCTGCCGAAACTGGGAGTGTCCAAATAGGCAGTGCCGAGCAGGACAGACCTCTCTGTAAG 3081

Qy        871 CAGGAGCCTGAGGCTAGTCAGGGACCAGTGTCCCAGTCAGAGACAGATGACATAGAAAAT 930

Db       3082 CAGGAGCCTGAGGCTAGTCAGGGACCAGTGTCCCAGTCAGAGACAGATGACATAGAAAAT 3141

Qy        931 CAGTTGGAAGGGAAGGCGACTCAGGGAAATACAGAGGAAAACCCCAGGGAAGCGGGCAAA 990

Db       3142 CAGTTGGAAGGGAAGGCGACTCAGGGAAATACAGAGGAAAACCCCAGGGAAGCGGGCAAA 3201

Qy        991 CCAAAGCCTGAGTGTGAGGTTTACGACCCCAATGCCCTGTACTGCATCTGCCGCCAGCCT 1050

Db       3202 CCAAAGCCTGAGTGTGAGGTTTACGACCCCAATGCCCTGTACTGCATCTGCCGCCAGCCT 3261

Qy       1051 CACAACAACAGGTTTATG 1068

Db       3262 CACAACAACAGGTAGGTG 3279


RESULT 7
BU522670
LOCUS       BU522670                874 bp    mRNA    linear   EST 13-SEP-2002
DEFINITION  AGENCOURT_10158094 NCI_CGAP_Co24 Mus musculus cDNA clone
            IMAGE:6529226 5', mRNA sequence.
ACCESSION   BU522670
VERSION     BU522670.1  GI:22830196
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 874)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14129  row: a  column: 02
            High quality sequence stop: 720.
FEATURES             Location/Qualifiers
     source          1. .874
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6529226"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Co24"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
ORIGIN

  Query Match           27.2%;  Score 780.8;  DB 13;  Length 874;
  Best Local Similarity   99.5%;  Pred. No. 6e-180;
  Matches  804;  Conservative    0;  Mismatches    2;  Indels    2;  Gaps    2;

Qy        797 GCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGCAGTGCCGAGCAGGACA 856

Db          1 GCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGCAGTGCCGAGCAGGACA 60

Qy        857 GACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTGTCCCAGTCAGAGACAG 916

Db         61 GACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTGTCCCAGTCAGAGACAG 120

Qy        917 ATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAATACAGAGGAAAACCCCA 976

Db        121 ATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAATACAGAGGAAAACCCCA 180

Qy        977 GGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCCAATGCCCTGTACTGCA 1036

Db        181 GGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCCAATGCCCTGTACTGCA 240

Qy       1037 TCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGATCGGTGTGAGGAGTGGT 1096

Db        241 TCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGATCGGTGTGAGGAGTGGT 300

Qy       1097 TCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAACGGGG 1156

Db        301 TCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAACGGGG 360

Qy       1157 AAGACTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGATGAGACAAACGGTAGCG 1216

Db        361 AAGACTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGATGAGACAAACGGTAGCG 420

Qy       1217 CCACCAATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGATGGCACAGACTGCACAA 1276

Db        421 CCACCGATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGATGGCACAGACTGCACAA 480

Qy       1277 GCATAGGGACAGTAGAGCAGAAGTCCGGAGAAGACCAGGGCATAAAGGGTAGGATTGAGA 1336

Db        481 GCATAGGGACAGTAGAGCAGAAGTCCGGAGAAGACCAGGGCATAAAGGGTAGGATTGAGA 540

Qy       1337 AGGCAGCAAACCCCAGCGGCAAGAAAAAACTCAAGATATTCCAGCCTGTCGTAGAGGCTC 1396

Db        541 AGGCAGCAAACCCCAGCGGCAAGAAAAAACTCAAGATATTCCAGCCTGTCGTAGAGGCTC 600
```

```
Qy       1397 CTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCACAGCCTGACTCTGTGT 1456
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCACAGCCTGACTCTGTGT 660

Qy       1457 ATTGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATGAGATTTCTAAGTTCAG 1516
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 ATTGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATGAGATTTCTAAGTTCAG 720

Qy       1517 GTAAAGAAC-AAAAAACAAAACCCAAGGAAAAGGTCAAGACGAAGCCAGAAAAGTTCAGT 1575
             ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 GTAAAGAACAAAAAAACAAAACCCAAGGAAAAGGTCAAGACGAAGCCAGAAAAGTTCAGT 780

Qy       1576 CTTCCAAAATGC-AGTGTTCAGGTGGGG 1602
             |||||||||||| |||||||||| ||||
Db         781 CTTCCAAAATGCAAGTGTTCAGGGGGGG 808


RESULT 8
BI078113
LOCUS       BI078113                796 bp    mRNA    linear   EST 20-JUN-2001
DEFINITION  602872420F1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5004317 5',
            mRNA sequence.
ACCESSION   BI078113
VERSION     BI078113.1  GI:14496443
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 796)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11042  row: g  column: 06
            High quality sequence stop: 742.
FEATURES             Location/Qualifiers
     source          1. .796
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N-3"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5004317"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam2"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
ORIGIN

  Query Match             24.5%;  Score 702.2;  DB 12;  Length 796;
  Best Local Similarity   98.1%;  Pred. No. 1e-160;
  Matches  784;  Conservative    0;  Mismatches    8;  Indels    7;  Gaps    7;

Qy       1439 CACAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCA 1498
             |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db          1 CACAGCCTGACTCTGTGTATAGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCA 60

Qy       1499 TGAGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAACCCAAGGAAAAGGTCAAGACGA 1558
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGAGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAACCCAAGGAAAAGGTCAAGACGA 120

Qy       1559 AGCCAGAA-AAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTGGGGATTAAAATCTCTTCT 1617
             |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 AGCCAGAACAAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTGGGGATTAAAATCTCTTCT 180

Qy       1618 GTGCACAAGAGACTAGCGTCAGAGAAAAGGGAAAACCCAGTGAAGAAAGTGATGCTGGCT 1677
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 GTGCACAAGAGACTAGCGTCAGAGAAAAGGGAAAACCCAGTGAAGAAAGTGATGCTGGCT 240

Qy       1678 TCCAGGAGTGAGACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGCACACCATCCTGGGCA 1737
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 TCCAGGAGTGAGACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGCACACCATCCTGGGCA 300

Qy       1738 AGTGACCACAACTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAGCCCACTGCACTCTCG 1797
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 AGTGACCACAACTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAGCCCACTGCACTCTCG 360

Qy       1798 CCCACCCTATTGAGTAAATGTACGTATCACCCAAAGGCTGGCTTCCCAGGCCCCTCCCAT 1857
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 CCCACCCTATTGAGTAAATGTACGTATCACCCAAAGGCTGGCTTCCCAGGCCCCTCCCAT 420

Qy       1858 CATCTGGGTGGCTGCCTGGGGCTGTCTAGGACCAGAGTCCTGGGTGTTCTGGTGCTGATA 1917
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 CATCTGGGTGGCTGCCTGGGGCTGTCTAGGACCAGAGTCCTGGGTGTTCTGGTGCTGATA 480

Qy       1918 GTAGCCAGCAGCTCACTGCCAGCCAGAAGCAGATACCAAGATGCCTCTGGA-CCCCAGGT 1976
             |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db         481 GTAGCCAGCAGCTCACTGCCAGCCAGAAGCAGATACCAAGATGCCTCTGGACCCCCAGGT 540

Qy       1977 GTTCCTGCCTAGCCTGTGGAGCCTCTCTGGGTGGTTCCTAAAGAGCTGTGTAGGCCTCAT 2036
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 GTTCCTGCCTAGCCTGTGGAGCCTCTCTGGGTGGTTCCTAAAGAGCTGTGTAGGCCTCAT 600

Qy       2037 GTTGGAGGCAATTTCTTATTTCAGTTTTAGGCCTTGGTAAGTATTTGAAC-TGCTCTC-A 2094
             |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||| |
Db         601 GTTGGAGGCAATTTCTTATTTCAGTTTTAGGCCTTGGTAAGTATTTGAACATGCTCTCAA 660

Qy       2095 ACAAGATGGGACAACATCAGTGCACTGTTCAGACTCAGTTCAGACTTGAGTTCCTCACAG 2154
             |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db         661 ACAAGATGGGACAACATCAGTGCACTGTTCAGACTCAGTTCAAACTTGAGTTCCTCACAG 720

Qy       2155 GACAGCCCAGCACAAACCATGGAGTTTTCCCACGTTACCTATTTTTCTTCAACATGGTCA 2214
             ||||| |||||||||||||| |||||| || ||||||| || || ||||| |||||||||
Db         721 GACAG-CCAGCACAAACCAT-GAGTTTCCCAACGTTACATAATTATCTTC-ACATGGTCA 777

Qy       2215 CCACTGCTCTTCTATAGAA 2233
             |||||| ||||||||||||
Db         778 CCACTGGTCTTCTATAGAA 796


RESULT 9
BY713603
LOCUS       BY713603                982 bp    mRNA    linear   EST 17-DEC-2002
DEFINITION  BY713603 RIKEN full-length enriched, 18 days pregnant, placenta and
            extra embryonic tissue Mus musculus cDNA clone 3830408B01 5', mRNA
            sequence.
ACCESSION   BY713603
VERSION     BY713603.1  GI:27125612
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 982)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
```

Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

MEDLINE 22354683

PUBMED 12466851

COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES Location/Qualifiers

source 1. .982
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="3830408B01"
/sex="female"
/tissue_type="placenta and extra embryonic tissue"
/dev_stage="18 days pregnant, adult"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 18 days pregnant, placenta and extra embryonic tissue"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 50.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I"

ORIGIN

Query Match 24.5%; Score 701; DB 13; Length 982;
Best Local Similarity 99.2%; Pred. No. 2.1e-160;
Matches 704; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
Qy          1 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         97 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 156

Qy         61 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        157 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 216

Qy        121 ACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        217 ACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTT 276

Qy        181 TATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        277 TATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGAT 336

Qy        241 AAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        337 AAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTC 396

Qy        301 AGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        397 AGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACG 456

Qy        361 GAGGCGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGA 420
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        457 GAGGTGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGA 516

Qy        421 CGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        517 CGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGG 576

Qy        481 AAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        577 AAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACT 636

Qy        541 GATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCT 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        637 GATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCTCT 696

Qy        601 GTATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAA 660
              |||||||||||||||| | ||||||||||||||||||||||||||||||||||||||||
Db        697 GTATCTGACTCCTTAGNGGAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAA 756

Qy        661 GAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGA 710
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        757 GAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCTTTACGTTGAGGGA 806
```

RESULT 10
CF169555
LOCUS CF169555 667 bp mRNA linear EST 25-JUL-2003
DEFINITION B0815C09-5 NIA Mouse Newborn Kidney cDNA Library (Long 1) Mus

musculus cDNA clone NIA:B0815C09 IMAGE:30468896 5', mRNA sequence.
ACCESSION CF169555
VERSION CF169555.1 GI:33279104
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 667)
AUTHORS Piao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.
TITLE Construction of long-transcript enriched cDNA libraries from submicrogram amounts of total RNAs by a universal PCR amplification method
JOURNAL Genome Res. 11 (9), 1553-1558 (2001)
MEDLINE 21429098
PUBMED 11544199
COMMENT Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
Plate: B0815 row: C column: 09
Seq primer: M13 Reverse
High quality sequence stop: 667
POLYA=No.
FEATURES Location/Qualifiers
source 1. .667
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="niaEST:B0815C09-5"
/db_xref="taxon:10090"
/clone="NIA:B0815C09 IMAGE:30468896"
/dev_stage="Newborn Kidney"
/lab_host="DH10B"
/clone_lib="NIA Mouse Newborn Kidney cDNA Library (Long 1)"
/note="Vector: pCMV-SPORT6 (Invitrogen); Site_1: SalI; Site_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). In brief, double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen: 5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTT-3'] from 26 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Lone-linker LL-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 3.0 kb. The library was constructed by Yulan Piao."
ORIGIN

Query Match 23.2%; Score 665.4; DB 14; Length 667;
Best Local Similarity 99.9%; Pred. No. 1e-151;
Matches 666; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy        628 CCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAAGAGGAAGAAGACACCTCTGACAGTGAC 687
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAAGAGGAAGAAGACACCTCTGACAGTGAC 60

Qy        688 AGTGATGGCCTTACGTTGAAGGAACTTCAGAACCGCCTTCGGAGAAAGCGAGAGCAAGAA 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGTGATGGCCTTACGTTGAAGGAACTTCAGAACCGCCTTCGGAGAAAGCGAGAGCAAGAA 120

Qy        748 CCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAA 807
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAA 180

Qy        808 GATTCTGCCGAAACTGGGAGTGTCCAAATAGGCAGTGCCGAGCAGGACAGACCTCTCTGT 867
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTCTGCCGAAACTGGGAGTGTCCAAATAGGCAGTGCCGAGCAGGACAGACCTCTCTGT 240

Qy        868 AAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTGTCCCAGTCAGAGACAGATGACATAGAA 927
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTGTCCCAGTCAGAGACAGATGACATAGAA 300

Qy        928 AATCAGTTGGAAGGGAAGGCGACTCAGGGAAATACAGAGGAAAACCCCAGGGAAGCGGGC 987
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AATCAGTTGGAAGGGAAGGCGACTCAGGGAAATACAGAGGAAAACCCCAGGGAAGCGGGC 360

Qy        988 AAACCAAAGCCTGAGTGTGAGGTTTACGACCCCAATGCCCTGTACTGCATCTGCCGCCAG 1047
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAACCAAAGCCTGAGTGTGAGGTTTACGACCCCAATGCCCTGTACTGCATCTGCCGCCAG 420

Qy       1048 CCTCACAACAACAGGTTTATGATCTGCTGTGATCGGTGTGAGGAGTGGTTCCATGGTGAC 1107
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCTCACAACAACAGGTTTATGATCTGCTGTGATCGGTGTGAGGAGTGGTTCCATGGTGAC 480

Qy       1108 TGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAACGGGGAAGACTACATC 1167
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAACGGGGAAGACTACATC 540

Qy       1168 TGCCCAAATTGCACCATTTTGCAAGTGCAGGATGAGACAAACGGTAGCGCCACCAATGAG 1227
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        541 TGCCCAAATTGCACCATTTTGCAAGTGCAGGATGAGACAAACGGTAGCGCCACCGATGAG 600

Qy       1228 CAGGACTCTGGGTGCAGATCTGTGGGTGCTGATGGCACAGACTGCACAAGCATAGGGACA 1287
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAGGACTCTGGGTGCAGATCTGTGGGTGCTGATGGCACAGACTGCACAAGCATAGGGACA 660

Qy       1288 GTAGAGC 1294
              |||||||
Db        661 GTAGAGC 667
```

RESULT 11
BB637982
LOCUS BB637982 691 bp mRNA linear EST 26-OCT-2001
DEFINITION BB637982 RIKEN full-length enriched, 3 days neonate thymus Mus musculus cDNA clone A630095A07 5', mRNA sequence.
ACCESSION BB637982
VERSION BB637982.1 GI:16473713
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 691)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

```
FEATURES             Location/Qualifiers
     source          1. .691
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="A630095A07"
                     /tissue_type="thymus"
                     /dev_stage="3 days neonate"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, 3 days neonate thymus"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda PLC I."
ORIGIN
```

```
  Query Match             21.7%;  Score 621.4;  DB 10;  Length 691;
  Best Local Similarity   99.8%;  Pred. No. 6.2e-141;
  Matches  622;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         68 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 127

Qy         61 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        128 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 187

Qy        121 ACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        188 ACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTT 247

Qy        181 TATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        248 TATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGAT 307

Qy        241 AAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        308 AAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTC 367

Qy        301 AGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        368 AGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACG 427

Qy        361 GAGGCGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGA 420
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        428 GAGGTGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGA 487

Qy        421 CGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        488 CGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGG 547

Qy        481 AAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        548 AAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACT 607

Qy        541 GATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        608 GATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCT 667

Qy        601 GTATCTGACTCCTTAGGGAAAGA 623
              |||||||||||||||||||||||
Db        668 GTATCTGACTCCTTAGGGAAAGA 690
```

```
RESULT 12
BQ771134
LOCUS       BQ771134                632 bp    mRNA    linear   EST 26-JUL-2002
DEFINITION  UI-M-FI0-byx-k-02-0-UI.r1 NIH_BMAP_FI0 Mus musculus cDNA clone
            IMAGE:6400633 5', mRNA sequence.
ACCESSION   BQ771134
VERSION     BQ771134.1  GI:21979610
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 632)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1. .632
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6400633"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 12.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FI0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel.First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
```

size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CAGCCACGAC. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN

Query Match 21.5%; Score 617.4; DB 13; Length 632;
Best Local Similarity 99.4%; Pred. No. 5.9e-140;
Matches 629; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```
Qy          13 CTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGGCAGTCCCCACTG 72
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 CTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGGCAGTCCCCACTG 60

Qy          73 TCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTGACTGACAACAGT 132
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 TCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTGACTGACAACAGT 120

Qy         133 GGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTTTATGTAGGATTT 192
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 GGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTTTATGTAGGATTT 180

Qy         193 CCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGATAAAGGGCACCTG 252
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 CCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGATAAAGGGCACCTG 240

Qy         253 AGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTCAGGAAAACCTGG 312
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 AGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTCAGGAAAACCTGG 300

Qy         313 GGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACGGAGGCGGACCCC 372
               |||||||||||||||||||| |||||||||||||||||||||||||||| |||||||
Db         301 GGTTTTCGAAGAACCACGATNGCCAAACGTGAGGGTGCAGGAGACACGGAGGTGGACCCC 360

Qy         373 AGTGAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGACGGCAACCAAAA 432
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 AGTGAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGACGGCAACCAAAA 420

Qy         433 CGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGGAAAAAGAATGTG 492
               |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         421 CGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCNAGGGAAAAAGAATGTG 480

Qy         493 CCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACTGATGTGGAGACA 552
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 CCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACTGATGTGGAGACA 540

Qy         553 GCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCTGTATCTGACTCC 612
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 GCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCTGTATCTGACTCC 600

Qy         613 TTAGGGAAAGAACATCCTGCCTCTTCTGAAAAG 645
               |||||||||||| ||||||||||||||||||||
Db         601 TTAGGGAAAGAA-ATCCTGCCTCTTCTGAAAAG 632
```

RESULT 13
CF535718
LOCUS CF535718 623 bp mRNA linear EST 12-SEP-2003
DEFINITION UI-M-GH0-cgz-h-20-0-UI.r1 NIH_BMAP_GH0 Mus musculus cDNA clone IMAGE:30534619 5', mRNA sequence.
ACCESSION CF535718
VERSION CF535718.1 GI:34587686
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 623)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at http://genome.uiowa.edu/distribution/mousefl.html
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
Seq primer: pYX-5.
FEATURES Location/Qualifiers
source 1. .623
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30534619"
/tissue_type="Whole brain"
/dev_stage="1, 5, and 15 days newborn"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_GH0"
/note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I; Site_2: Not I; The library was constructed according Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGAACTGAAT. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN

Query Match 21.2%; Score 607.4; DB 14; Length 623;
Best Local Similarity 99.4%; Pred. No. 1.7e-137;
Matches 619; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```
Qy         302 GGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACGG 361
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 GGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACGG 60

Qy         362 AGGCGGACCCCAGT-GAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGA 420
               ||| |||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db          61 AGGTGGACCCCAGTNGAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGA 120

Qy         421 CGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGG 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 CGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGG 180

Qy         481 AAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACT 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 AAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACT 240

Qy         541 GATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCT 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 GATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCT 300

Qy         601 GTATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAA 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 GTATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAA 360

Qy         661 GAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAAC 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 GAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAAC 420
```

```
Qy        721 CGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAAT 480

Qy        781 CGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGC 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        481 CGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCANATAGGC 540

Qy        841 AGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTG 600

Qy        901 TCCCAGTCAGAGACAGATGACAT 923
              || ||||||||||||||||||||
Db        601 TCNCAGTCAGAGACAGATGACAT 623
```

RESULT 14
BB636983

```
LOCUS       BB636983                645 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB636983 RIKEN full-length enriched, adult male aorta and vein Mus
            musculus cDNA clone A530047I15 5', mRNA sequence.
ACCESSION   BB636983
VERSION     BB636983.1  GI:16472769
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 645)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
            Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
            Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
            Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
            and Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
            Sugahara,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I.,
            Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
            e mouse tissues.
FEATURES            Location/Qualifiers
     source         1. .645
                    /organism="Mus musculus"
                    /mol_type="mRNA"
                    /db_xref="taxon:10090"
                    /clone="A530047I15"
                    /sex="male"
                    /tissue_type="aorta and vein"
                    /dev_stage="adult"
                    /lab_host="DH10B"
                    /clone_lib="RIKEN full-length enriched, adult male aorta
                    and vein"
                    /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                    prepared and sequenced in Mouse Genome Encyclopedia
                    Project of Genome Exploration Research Group in Riken
                    Genomic Sciences Center and Genome Science Laboratory in
                    RIKEN. Division of Experimental Animal Research in Riken
                    contributed to prepare mouse tissues. 1st strand cDNA was
                    primed with a primer [5'
                    GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                    prepared by using trehalose thermo-activated reverse
                    transcriptase and subsequently enriched for full-length by
                    cap-trapper. cDNA went through one round of normalization
                    to Rot = 20.0 and subtraction to Rot = 459.0. Second
                    strand cDNA was prepared with the primer adapter of
                    sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                    3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                    modified pBluescript KS(+) after bulk excision from Lambda
                    FLC I."
ORIGIN
```

```
  Query Match             20.7%;  Score 594.6;  DB 10;  Length 645;
  Best Local Similarity   99.2%;  Pred. No. 2.3e-134;
  Matches  608;  Conservative    0;  Mismatches    4;  Indels    1;  Gaps    1;

Qy          1 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         34 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 93

Qy         61 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         94 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 153

Qy        121 ACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        154 ACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTT 213

Qy        181 TATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        214 TATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGAT 273

Qy        241 AAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        274 AAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTC 333

Qy        301 AGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        334 AGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACG 393

Qy        361 GAGGCGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGA 420
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        394 GAGGTGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGA 453

Qy        421 CGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        454 CGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGG 513

Qy        481 AAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACT 540
              |||||||||||||||||||||||||||||||||||||||||||||| || ||||||||||
Db        514 AAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACAT-TTTCACAGTCACT 572

Qy        541 GATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCT 600
```

```
                    |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        573 GATGTGGAGACAGCTTCCGAAGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCT 632

Qy        601 GTATCTGACTCCT 613
              |||| ||||||||
Db        633 GTATTTGACTCCT 645
```

RESULT 15
BY752619
LOCUS BY752619 746 bp mRNA linear EST 17-DEC-2002
DEFINITION BY752619 RIKEN full-length enriched, 9.5 days embryo parthenogenote Mus musculus cDNA clone B130065J04 5', mRNA sequence.
ACCESSION BY752619
VERSION BY752619.1 GI:27183600
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 746)
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
MEDLINE 22354683
PUBMED 12466851
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by Dr. Tomohiro Kono (Department of Animal Science, Tokyo University of Agriculture, 1737 Hunako Atsugi City, Kanagawa Prefecture, Japan) whose assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .746
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="B130065J04"
/tissue_type="parthenogenote"
/dev_stage="9.5 days embryo"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 9.5 days embryo parthenogenote"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of subtraction to Rot = 229.0 Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda PLC I."
ORIGIN

Query Match 20.6%; Score 591.6; DB 13; Length 746;
Best Local Similarity 92.7%; Pred. No. 1.2e-133;
Matches 651; Conservative 0; Mismatches 48; Indels 3; Gaps 3;

```
Qy          1 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         17 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 76

Qy         61 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         77 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 136

Qy        121 ACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        137 ACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTT 196

Qy        181 TATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        197 TATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGAT 256

Qy        241 AAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        257 AAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTC 316

Qy        301 AGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACG 360
```

```
Db        317 AGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACG 376

Qy        361 GAGGCGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGA 420

Db        377 GAGGTGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGA 436

Qy        421 CGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGG 480

Db        437 CGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCCAGGG 496

Qy        481 AAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACT 540

Db        497 AAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACCTTTTCCACAGTCACT 556

Qy        541 GATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCT 600

Db        557 GATGTGGAGACGGCTTCCGAGGGGAGCG-TGAAAGCAGTTCTGAAATCAAAAGGGGCCCT 615

Qy        601 GTATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAA 660

Db        616 GT-TCTTACTCCTTAGGGGAA-AAAATCCTGCCTTTTTGAAAAGGCAAAGGAGGGGAAAA 673

Qy        661 GAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACG 702

Db        674 GAANNAAAAACCTTTNAAGNGAACGGAAGGCCTTACCTTAAG 715
```

Search completed: April 23, 2004, 18:03:08
Job time : 7086.26 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 23, 2004, 07:35:03 ; Search time 174.89 Seconds
(without alignments)
8281.924 Million cell updates/sec

Title: US-09-787-016A-1
Perfect score: 2610
Sequence: 1 ctcggtggccgtccgcccac.....actcttaagatcatatcctg 2610

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	202.8	7.8	472	4	US-09-621-976-18485	Sequence 18485, A
c	2	67	2.6	7218	1	US-08-232-463-14	Sequence 14, Appl
	3	65	2.5	6256	4	US-09-620-312D-269	Sequence 269, App
	4	57.4	2.2	7218	1	US-08-232-463-14	Sequence 14, Appl
c	5	44.2	1.7	1926	4	US-09-249-585A-4	Sequence 4, Appli
c	6	44.2	1.7	1931	2	US-09-130-114-2	Sequence 2, Appli
c	7	42.6	1.6	16442	3	US-08-781-891-208	Sequence 208, App
c	8	42.6	1.6	16442	4	US-09-618-166-208	Sequence 208, App
	9	41.8	1.6	4951	2	US-08-867-030B-5	Sequence 5, Appli
	10	41.8	1.6	4951	5	PCT-US95-06119-5	Sequence 5, Appli
	11	41.4	1.6	1068	4	US-09-710-794-3	Sequence 3, Appli
	12	41	1.6	669	4	US-09-669-751-79	Sequence 79, Appl
	13	40.8	1.6	364	4	US-09-621-976-17202	Sequence 17202, A
	14	40.2	1.5	289	3	US-09-007-005-17	Sequence 17, Appl
	15	40.2	1.5	289	3	US-09-244-796-17	Sequence 17, Appl
	16	40.2	1.5	6232	3	US-08-456-200B-11	Sequence 11, Appl
c	17	40.2	1.5	51259	3	US-08-781-891-209	Sequence 209, App
c	18	40.2	1.5	51259	4	US-09-618-166-209	Sequence 209, App
	19	40	1.5	1926	4	US-09-249-585A-2	Sequence 2, Appli
	20	40	1.5	1926	4	US-09-410-399-3	Sequence 3, Appli
	21	40	1.5	2580	3	US-09-050-863-2	Sequence 2, Appli
	22	40	1.5	2580	4	US-09-359-081-2	Sequence 2, Appli
c	23	40	1.5	5452	2	US-09-130-114-1	Sequence 1, Appli
c	24	40	1.5	8705	4	US-09-647-344A-14	Sequence 14, Appl
	25	40	1.5	9600	3	US-08-910-647-1	Sequence 1, Appli
	26	40	1.5	9600	4	US-09-620-925-1	Sequence 1, Appli
	27	40	1.5	10596	1	US-07-884-811-15	Sequence 15, Appl
	28	40	1.5	10596	1	US-07-885-971-15	Sequence 15, Appl
	29	40	1.5	10596	1	US-08-087-783A-15	Sequence 15, Appl
	30	40	1.5	10596	1	US-08-194-088B-15	Sequence 15, Appl
	31	40	1.5	10596	2	US-08-194-087-15	Sequence 15, Appl
	32	40	1.5	10596	5	PCT-US93-04648-15	Sequence 15, Appl
c	33	40	1.5	16080	4	US-09-724-566A-48	Sequence 48, Appl
c	34	39.2	1.5	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	35	39.2	1.5	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	36	39	1.5	2338	1	US-08-425-069-1	Sequence 1, Appli
	37	39	1.5	2338	2	US-08-317-844B-1	Sequence 1, Appli
	38	39	1.5	2561	4	US-09-616-289-48	Sequence 48, Appl
c	39	38.8	1.5	1779	4	US-09-252-991A-3864	Sequence 3864, Ap
	40	38.8	1.5	1842	4	US-09-252-991A-3597	Sequence 3597, Ap
c	41	38.8	1.5	1923	4	US-09-252-991A-3772	Sequence 3772, Ap
	42	38.6	1.5	2846	3	US-09-613-182-5	Sequence 5, Appli
c	43	37.8	1.4	245	1	US-08-139-937-8	Sequence 8, Appli
c	44	37.8	1.4	245	5	PCT-US93-11310-8	Sequence 8, Appli
c	45	37.8	1.4	505	4	US-09-621-976-15639	Sequence 15639, A

ALIGNMENTS

RESULT 1
US-09-621-976-18485
; Sequence 18485, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 18485
; LENGTH: 472
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-18485

Query Match 7.8%; Score 202.8; DB 4; Length 472;
Best Local Similarity 99.0%; Pred. No. 8.7e-50;
Matches 204; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy         63 GCGGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCC 122
               | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 GAGGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCC 82

Qy        123 TGTTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGAC 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 TGTTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGAC 142

Qy        183 CCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCA 242
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 CCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCA 202

Qy        243 AGCTTTCAGTGTTGGAGCAGGTATGG 268
              |||||||||||||||||||||||| ||
Db        203 AGCTTTCAGTGTTGGAGCAGGTAAGG 228
```

RESULT 2
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIPLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

```
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match             2.6%;  Score 67;  DB 1;  Length 7218;
  Best Local Similarity   4.6%;  Pred. No. 8.4e-09;
  Matches   19;  Conservative  235;  Mismatches  155;  Indels    0;  Gaps    0;

Qy        654 AGCTGTGAAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGA 713
              ||  ||  ||  ||    |  ||  |  |  |  ::::: :::::::::::: ::  : :: ::
Db       1464 AGTAGTTAAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1405

Qy        714 CACCTCCGATAGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAG 773
              :    :: :: :: :: :: ::   ::    :::::::    ::::  :     : ::
Db       1404 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1345

Qy        774 GAAGCGGGAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAA 833
              :::: ::::: :::::   :  :::::::     ::::::::   ::::  :   : :::::
Db       1344 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1285

Qy        834 GAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGA 893
              :::: :  ::::::::::    :  ::::  : :::    ::::  :: :: :  : :::
Db       1284 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1225

Qy        894 GGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGG 953
              ::: :    :   :: ::: :::::   ::::: ::  :::::: : :    :::  ::
Db       1224 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1165

Qy        954 GAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGA 1013
              :::::: :: :::::::: :::  :::::::::::: ::    :::: :  ::::: :::::
Db       1164 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1105

Qy       1014 GCCTGGAGACTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCC 1062
              :   :::::   :::  ::  ::::   ::: : :::: |  | | | |
Db       1104 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRATCGCAAGCTC 1056
```

```
RESULT 3
US-09-620-312D-269
; Sequence 269, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Chen, Rui-hong
;  APPLICANT:  Zhao, Qing A.
;  APPLICANT:  Wehrman, Tom
;  APPLICANT:  Xue, Aidong J.
;  APPLICANT:  Yang, Yonghong
;  APPLICANT:  Wang, Jian-Rui
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Ma, Yunqing
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Wang, Zhiwei
;  APPLICANT:  John Tillinghast
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. 6569662el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 784CIP2B
;  CURRENT APPLICATION NUMBER: US/09/620,312D
;  CURRENT FILING DATE:  2000-07-19
;  PRIOR APPLICATION NUMBER: 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 1105
;  SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 269
;   LENGTH: 6256
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (257)..(5443)
US-09-620-312D-269

  Query Match             2.5%;  Score 65;  DB 4;  Length 6256;
  Best Local Similarity   62.7%;  Pred. No. 3e-08;
  Matches  101;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy       1060 CCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGT 1119
              ||||  |   | ||  | |  |||||    | ||  ||   ||||||||||||| || |||||
Db       1469 CCCAGCAAGCAGTGTGGGTTTTGCAAAAAACCACATGGCAACAGGTTTATGGTTGGCTGT 1528

Qy       1120 GACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGG 1179
              |   | |||||  || |||||||||||| ||||||||  ||   |    |||  ||    |
Db       1529 GGGAGATGTGATGACTGGTTTCATGGTGATTGTGTTGGGTTAAGTCTTTCTCAAGCACAG 1588

Qy       1180 CTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACTG 1220
              |   |||    | | |   |||| ||| ||||   ||| ||
Db       1589 CAGATGGGCGAGGAAGACAAAGAATATGTCTGTGTAAAATG 1629


RESULT 4
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
```

```
;       STATE:  VA
;       COUNTRY:  USA
;       ZIP:  22313-0299
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/232,463
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/935,313
;       FILING DATE:
;       APPLICATION NUMBER:  EP 91 114 300.6
;       FILING DATE:  26-AUG-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BENT, Stephen A.
;       REGISTRATION NUMBER:  29,768
;       REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703)836-9300
;       TELEFAX:  (703)683-4109
;       TELEX:  899149
;   INFORMATION FOR SEQ ID NO:  14:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  7218 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     IMMEDIATE SOURCE:
;       CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match             2.2%;  Score 57.4;  DB 1;  Length 7218;
  Best Local Similarity   3.4%;  Pred. No. 5.8e-06;
  Matches   13;  Conservative  222;  Mismatches  148;  Indels    0;  Gaps    0;

Qy       1970 GGCTGCACCATCAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTC 2029
Db       1041 GGCTGCAGGTCGAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1100

Qy       2030 CATCCCTAGTCTGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGTG 2089
Db       1101 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1160

Qy       2090 AACGTTCACCTTAATGATTGATTCTTTAATCTCTGTTTTCACTCTCAGGCTCTGGTAAGT 2149
Db       1161 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1220

Qy       2150 ATTTGTATTCTCTTCATCCCAGTCTGATTGCATAGCCACACTGCCCGGCACGCCACATCC 2209
Db       1221 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1280

Qy       2210 ACCCCTGTCTGCACATGAGTTGTTCTGACAACAGCGCTGTATACGCTTCAGTTTTTCCAC 2269
Db       1281 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1340

Qy       2270 ATTGTCCACGGCCAGCACATGAAAGCATCACTTCTTTTTTATGTTGTGGGAATCTTTGCA 2329
Db       1341 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1400

Qy       2330 AGTTAGTGTTGCATCTGATTTTC 2352
Db       1401 YYYYYYYYYYYYYYYYYYYYY 1423


RESULT 5
US-09-249-585A-4/c
; Sequence 4, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
;  APPLICANT: Horlick, Robert
;  TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
;  FILE REFERENCE: 0867/0D905
;  CURRENT APPLICATION NUMBER: US/09/249,585A
;  CURRENT FILING DATE:  1999-02-11
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
;   LENGTH: 1926
;   TYPE: DNA
;   ORGANISM: Epstein Barr Virus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(1926)
;   OTHER INFORMATION: template strand of EBNA-1 DNA
US-09-249-585A-4

  Query Match             1.7%;  Score 44.2;  DB 4;  Length 1926;
  Best Local Similarity   44.8%;  Pred. No. 0.019;
  Matches  169;  Conservative    0;  Mismatches  208;  Indels    0;  Gaps    0;

Qy        682 GAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGGCCTG 741
Db        886 GACGGGGAGGACGGGGAGGACGAGGACGGGGAGGACGGGGAGGAGGACGAGGACGGGGAG 827

Qy        742 ACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGAGG 801
Db        826 GACGGGGAGGAGGACGAGGACGGGGAGGACGGGGAGGACGAGGACGGGGAGGACGGGGAG 767

Qy        802 CCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCGAG 861
Db        766 GACGAGGACGGGGAGGACGGGGAGGACGAGGACGGGGAGGACGAGGACGGGGAGGACGAG 707

Qy        862 ACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGCCC 921
Db        706 GACGGGGAGGACGAGGACGGGGAGGACGGGGAGGACGGGGAGGAGGACGAGGACGGGGAG 647

Qy        922 GAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGGAG 981
Db        646 GACGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGAGGACGGGGAGGAGGACGAGGAC 587

Qy        982 GGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACTTGGGCCGACCGAAGCCT 1041
Db        586 GGGGAGGACGGGGAGGACGGGGAGGAGGACGAGGACGGGGAGGACGGGGAGGAGGACGAG 527

Qy       1042 GAATGTGAGGGTTACGA 1058
Db        526 GACGGGGAGGAGGACGA 510


RESULT 6
US-09-130-114-2/c
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
;  APPLICANT: Horlick, Robert A.
;  APPLICANT: Damaj, Bassam B.
;  APPLICANT: Robbins, Alan K.
;  TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
;  TITLE OF INVENTION: From Multiple Transfected Episomes
;  FILE REFERENCE: 0867/1D903US1
;  CURRENT APPLICATION NUMBER: US/09/130,114
;  CURRENT FILING DATE:  1998-08-06
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 1931
;   TYPE: DNA
;   ORGANISM: EBNA
US-09-130-114-2

  Query Match             1.7%;  Score 44.2;  DB 2;  Length 1931;
  Best Local Similarity   44.8%;  Pred. No. 0.019;
```

```
  Matches  169;  Conservative    0;  Mismatches  208;  Indels    0;  Gaps    0;

Qy        682 GAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGGCCTG 741
              ||   || | | || | ||| ||  |||  ||    |   ||    |||   || ||   |
Db        886 GACGGGGAGGACGGGGAGGACGAGGACGGGGAGGACGGGGAGGAGGACGAGGACGGGGAG 827

Qy        742 ACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGAGG 801
              |  | | |||    || | |      | || ||  ||  | | |     | | |
Db        826 GACGGGGAGGAGGACGAGGACGGGGAGGACGGGGAGGACGAGGACGGGGAGGACGGGGAG 767

Qy        802 CCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCGAG 861
              |  | | |||    | |      | |   |    |||||  ||    | ||||
Db        766 GACGAGGACGGGGAGGACGGGGAGGACGAGGACGGGGAGGACGAGGACGGGGAGGACGAG 707

Qy        862 ACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGCCC 921
              | ||    ||||| | | ||    | ||||| ||    |      || | | |
Db        706 GACGGGGAGGACGAGGACGGGGAGGACGGGGAGGACGGGGAGGAGGACGAGGACGGGGAG 647

Qy        922 GAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGGAG 981
              ||  | ||   || ||  | |    | | | | | || ||| | |||    |   |||
Db        646 GACGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGAGGACGGGGAGGAGGACGAGGAC 587

Qy        982 GGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACTTGGGCCGACCGAAGCCT 1041
              ||  |||  |   |||||    | || || |||   || ||   ||      | |
Db        586 GGGGAGGACGGGGAGGACGGGGAGGAGGACGAGGACGGGGAGGACGGGGAGGAGGACGAG 527

Qy       1042 GAATGTGAGGGTTACGA 1058
              ||  | ||||   ||||
Db        526 GACGGGGAGGAGGACGA 510


RESULT 7
US-08-781-891-208/c
; Sequence 208, Application US/08781891
; Patent No. 6090620
;  GENERAL INFORMATION:
;    APPLICANT:  Fu, Ying-Hui
;    APPLICANT:  Yu, Chang-En
;    APPLICANT:  Oshima, Junko
;    APPLICANT:  Mulligan, John T.
;    APPLICANT:  Schellenberg, Gerald D.
;    TITLE OF INVENTION:  GENE AND GENE PRODUCTS RELATED TO
;    TITLE OF INVENTION:  WERNER'S SYNDROME
;    NUMBER OF SEQUENCES:  209
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,891
;      FILING DATE:  27-DEC-1996
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6090620tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  240052.419
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  208:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  16442 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear

US-08-781-891-208

  Query Match             1.6%;  Score 42.6;  DB 3;  Length 16442;
  Best Local Similarity   48.5%;  Pred. No. 0.24;
  Matches  117;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy        774 GAAGCGGGAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAA 833
              | ||| ||| ||||||      |||  |   | | |   |    ||         || | |
Db      16430 GGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGCAGGAGCAGGAGGAGGAGGAGGAGCAGGA 16371

Qy        834 GAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGA 893
              | ||   | |||||||  |  |   ||  | |      ||||     ||    | |||
Db      16370 GCAGGAGCAGGAGGAGGAGCAGGAGCAGGAGGAGCAGGAGGAGGAGGAGGAGCAGGAGGA 16311

Qy        894 GGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGG 953
              |   |   |      |||||||  ||| | || |||| ||  | |   |||| |
Db      16310 GCAGGAGGAGGAGGAGGAGCAGGAGGAGGAGCAGGAGCAGGAGGAGGAGGAGCAGGAGGA 16251

Qy        954 GAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGA 1013
              | | || ||   |||    ||  |||||  ||| ||   ||||      | | ||| |
Db      16250 GCAGGAGGAGGAGGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGTAGAAGGAGAA 16191

Qy       1014 G 1014
              |
Db      16190 G 16190


RESULT 8
US-09-618-166-208/c
; Sequence 208, Application US/09618166
; Patent No. 6583112
;   GENERAL INFORMATION:
;        APPLICANT: Fu, Ying-Hui
;                   Yu, Chang-En
;                   Oshima, Junko
;                   Mulligan, John T.
;                   Schellenberg, Gerald D.
;        TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
;                            WERNER'S SYNDROME
;        NUMBER OF SEQUENCES: 209
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Seed Intellectual Property Law Group
;             STREET: 701 Fifth Avenue, Suite 6300
;             CITY: Seattle
;             STATE: Washington
;             COUNTRY: USA
;             ZIP: 98104-7092
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/618,166
;             FILING DATE: 17-Jul-2000
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McMasters, David D.
;             REGISTRATION NUMBER: 33,963
;             REFERENCE/DOCKET NUMBER: 240052.419C1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (206) 622-4900
;             TELEFAX: (206) 682-6031
;   INFORMATION FOR SEQ ID NO: 208:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 16442 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 208:
US-09-618-166-208

  Query Match             1.6%;  Score 42.6;  DB 4;  Length 16442;
```

```
  Best Local Similarity   48.5%;  Pred. No. 0.24;
  Matches  117;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy        774 GAAGCGGGAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAA 833
              | ||| ||| |||||||      ||| |   | | |    |     ||          || | |
Db      16430 GGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGCAGGAGCAGGAGGAGGAGGAGGAGCAGGA 16371

Qy        834 GAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGA 893
              | ||    | |||||||| |  |    ||   | |       ||||       ||     | |||
Db      16370 GCAGGAGCAGGAGGAGGAGCAGGAGCAGGAGGAGCAGGAGGAGGAGGAGGAGCAGGAGGA 16311

Qy        894 GGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGG 953
              |   |    |       ||||||||   ||| | || |||| ||  | |   |||| |
Db      16310 GCAGGAGGAGGAGGAGGAGCAGGAGGAGGAGCAGGAGCAGGAGGAGGAGGAGCAGGAGGA 16251

Qy        954 GAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGA 1013
              | | || ||    |||    ||  |||||   ||| ||   ||||      | | ||| |
Db      16250 GCAGGAGGAGGAGGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGTAGAAGGAGAA 16191

Qy       1014 G 1014
              |
Db      16190 G 16190


RESULT 9
US-08-867-030B-5
; Sequence 5, Application US/08867030B
; Patent No. 5948900
;  GENERAL INFORMATION:
;    APPLICANT:  Yother et al.
;    TITLE OF INVENTION:  Streptococcus pneumoniae
;    TITLE OF INVENTION:  Capsular Polysaccharide Genes and Flanking Regions
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Benjamin Aaron Adler, Ph.D., J.D.
;      STREET:  8011 Candle Lane
;      CITY:  Houston
;      STATE:  TX
;      COUNTRY:  USA
;      ZIP:  77071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch floppy
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh
;      SOFTWARE:  Microsoft Word for Macintosh
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/867,030B
;      FILING DATE:  June 2, 1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/243,546
;      FILING DATE:  May 16, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Benjamin Aaron Adler, Ph.D., J.D.
;      REGISTRATION NUMBER:  35,423
;      REFERENCE/DOCKET NUMBER:  D5923
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (713) 777-2321
;      TELEFAX:  (713) 777-6908
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4951 bp
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
;      DESCRIPTION:  other nucleic acid
;    HYPOTHETICAL:  no
;    ANTI-SENSE:  no
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
;    IMMEDIATE SOURCE:
;    POSITION IN GENOME:

US-08-867-030B-5

  Query Match             1.6%;  Score 41.8;  DB 2;  Length 4951;
  Best Local Similarity   49.8%;  Pred. No. 0.19;
  Matches  106;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy       2237 ACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGGCCAGCACATGAAAGCA 2296
              | || | | || || ||||       ||| || | |    || |  | | | | | | |
Db       1238 AGAAAAACTCTATACACGCGATTTATTTGGCAGAGAATAAGGGGAAATAATTTTTATGTA 1297

Qy       2297 TCACTTCTTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTGCATCTGATTTTCAGGT 2356
              |   ||  |||| |||||||| ||| | ||| ||   |  || |    |  ||| | |
Db       1298 TACATTTATTTTAATGTTGTTGGATTTTTTTCAGAATCATGATTTTCATTTCTTTATGTT 1357

Qy       2357 GTACATTTATTTTTGACTGGGCAGATAGGGGATTTTTTTTTTTTCCATGTCCGATTCAC 2416
              ||   ||  |||   ||      | | || | || | | | ||||     |||| ||| |
Db       1358 GTTTTTTGTCTTTATTCTTATTCGTTGGGCGGTTATATATTTTCATGCTGTCAGATATAA 1417

Qy       2417 ACGCTACACACCCACATGAACACATTCGAACTT 2449
              |||||     |    ||  ||  ||| ||
Db       1418 GTCCTACAGTTGTAGTGTAAGTGATGAGAAGTT 1450


RESULT 10
PCT-US95-06119-5
; Sequence 5, Application PC/TUS9506119
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  STREPTOCOCCUS PNEUMONIAE CAPSULAR
;    TITLE OF INVENTION:  POLYSACCHARIDE GENES AND FLANKING REGIONS
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  TX
;      COUNTRY:  United States of America
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS/ASCII
;      SOFTWARE:  PatentIn Release #1.0, Version
;      SOFTWARE:  #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/06119
;      FILING DATE:  CONCURRENTLY HEREWITH
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/243,546
;      FILING DATE:  16-MAY-1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  AMCY018P--
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (713) 789-2679
;      TELEX:  79-0924
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4951 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "DNA"
PCT-US95-06119-5

  Query Match             1.6%;  Score 41.8;  DB 5;  Length 4951;
  Best Local Similarity   49.8%;  Pred. No. 0.19;
  Matches  106;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;
```

```
Qy       2237 ACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGGCCAGCACATGAAAGCA 2296
Db       1238 AGAAAAACTCTATACACGCGATTTATTTGGCAGAGAATAAGGGGAAATAATTTTTATGTA 1297

Qy       2297 TCACTTCTTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTGCATCTGATTTTCAGGT 2356
Db       1298 TACATTTATTTTAATGTTGTTGGATTTTTTTCAGAATCATGATTTTCATTTCTTTATGTT 1357

Qy       2357 GTACATTTATTTTTGACTGGGCAGATAGGGGATTTTTTTTTTTTTCCATGTCCGATTCAC 2416
Db       1358 GTTTTTTGTCTTTATTCTTATTCGTTGGGCGGTTATATATTTTCATGCTGTCAGATATAA 1417

Qy       2417 ACGCTACACACCCACATGAACACATTCGAACTT 2449
Db       1418 GTCCTACAGTTGTAGTGTAAGTGATGAGAAGTT 1450
```

RESULT 11
US-09-710-794-3
; Sequence 3, Application US/09710794
; Patent No. 6573069
; GENERAL INFORMATION:
; APPLICANT: Holloway, James L.
; APPLICANT: Gao, Zeren
; APPLICANT: Whitmore, Theodore E.
; TITLE OF INVENTION: NOVEL CRIB PROTEIN ZMSE1
; FILE REFERENCE: 99-76
; CURRENT APPLICATION NUMBER: US/09/710,794
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/164,685
; PRIOR FILING DATE: 1999-11-10
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1068
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate polynucleotide seuence for human zmse1
; NAME/KEY: misc_feature
; LOCATION: (1)...(1068)
; OTHER INFORMATION: n = A,T,C or G
US-09-710-794-3

Query Match 1.6%; Score 41.4; DB 4; Length 1068;
Best Local Similarity 22.4%; Pred. No. 0.086;
Matches 102; Conservative 79; Mismatches 275; Indels 0; Gaps 0;

```
Qy        513 CCGCGGCAGGAGGAGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGCCCACGTCCTGCCC 572
Db        102 YMGNCAYACNATGCAYGTNGGNMGNGCNGGNGAYGCNTTYGGNGAYACNWSNTTYYTNAA 161

Qy        573 CGCCACAGACGCCGAGACAGCCTCCGAGGGCAGCGTGGAAAGCGCTTCTGAGACCAGAAG 632
Db        162 YWSNAARGCNGGNGARCCNGAYGGNGARWSNYTNGAYGARCARCCNWSNWSNWSNWSNWS 221

Qy        633 CGGCCCCCAGTCTGCTTCCACAGCTGTGAAGGAACGACCAGCCTCTTCTGAAAAGGTGAA 692
Db        222 NAARMGNWSNYTNYTNWSNMGNAARTTYMGNGGNWSNAARMGNWSNCARWSNGTNACNMG 281

Qy        693 AGGAGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGGCCTGACCTTGAAAGA 752
Db        282 NGGNGARMGNGARCARMGNGAYATGYTNGGNWSNYTNMGNGAYWSNGCNYTNTTYGTNAA 341

Qy        753 GCTTCAGAATCGCCTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGAGGCCCCTGAAAGG 812
Db        342 RAAYGCNATGWSNYTNCCNCARYTNAAYGARAARGARGCNGCNGARAARGGNACNWSNAA 401

Qy        813 GATCCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTGGGCTC 872
Db        402 RYTNCCNAARWSNYTNWSNWSNWSNCCNGTNAARAARGCNAAYGAYGGNGARGGNGGNGA 461
```

```
Qy        873 CGAGGCCAGTGACACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAACGATCA 932
Db        462 YGARGARGCNGGNACNGARGARGCNGTNCCNMGNMGNAAYGGNGCNGCNGGNCCNCAYWS 521

Qy        933 GGGGGTTGTGTCCCAGGCTGGGAAAGATGACAGAGA 968
Db        522 NCCNGAYCCNYTNYTNGAYGARCARGCNTTYGGNGA 557
```

RESULT 12
US-09-669-751-79
; Sequence 79, Application US/09669751
; Patent No. 6551575
; GENERAL INFORMATION:
; APPLICANT: Greenspan, Ralph J.
; TITLE OF INVENTION: Methods for Identifying Compounds for
; TITLE OF INVENTION: Motion Sickness, Vertigo and Other Disorders Related to
; TITLE OF INVENTION: Balance and the Perception of Gravity
; FILE REFERENCE: P-NI 3864
; CURRENT APPLICATION NUMBER: US/09/669,751
; CURRENT FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US 60/168,579
; PRIOR FILING DATE: 1999-12-02
; NUMBER OF SEQ ID NOS: 261
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 79
; LENGTH: 669
; TYPE: DNA
; ORGANISM: Drosophila
US-09-669-751-79

Query Match 1.6%; Score 41; DB 4; Length 669;
Best Local Similarity 46.1%; Pred. No. 0.082;
Matches 137; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

```
Qy        742 ACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGAGG 801
Db        201 ACCTTGTGGCAGAAATTGAACTGCGTTCCGTGCAACCGTTTTGTGTTCTTCATGATCGGG 260

Qy        802 CCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCGAG 861
Db        261 GCGGGCATCGGATTATTCTGGGACCATATGAAGAAGGAGGCCCAGAAGGCGGCCGAGGAG 320

Qy        862 ACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGCCC 921
Db        321 GCCGCGAAATCGCCCAAGGAGAAGGAGAAGGAGGCCAAGGAGCGCGAGAAGAAGCAGAAG 380

Qy        922 GAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGGAG 981
Db        381 GAAAAAGAGAAGGCCGAAAAGGAGAAGGCGGAGAAGGAGAAGAAGGAGAGGGAGAAGAAG 440

Qy        982 GGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACTTGGGCCGACCGAAG 1038
Db        441 GAGAAGGAAAAGAAGGAGAAGGAAAAGAAGGAGAAGGAGGCCAAGGGCAAAGAGAAG 497
```

RESULT 13
US-09-621-976-17202
; Sequence 17202, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 17202
; LENGTH: 364
; TYPE: DNA

```
;   ORGANISM: Homo sapiens
US-09-621-976-17202

  Query Match             1.6%;  Score 40.8;  DB 4;  Length 364;
  Best Local Similarity   13.1%;  Pred. No. 0.062;
  Matches   36;  Conservative  123;  Mismatches  115;  Indels    0;  Gaps    0;

Qy        426 GCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCGT 485
              | |:| || :|| :: :|  : :  :  :  |::| :  :|: :: :      :: :
Db         10 GRARCCGCCKGGAGYSGMCKSSRSYGRRSSCCGSMGWSGCSCSKRSWSRCRCMKSMWSWM 69

Qy        486 GGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGGA 545
              :::::: :| |:| ::| ::  : | : | : : : : |  :::: :::||::: ::
Db         70 MYMRSMKYKRSTCASCKYKGGKMACMTCWSTGAMYRYMASYGWCYSYMARYYTCYSKYRM 129

Qy        546 GGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCAG 605
              :  : : : :  ||::|  :   :: :  ::: :  :  :: :  ||:   : :|:::
Db        130 WKYCYRKYRSRGMCCMWCAGSGMCYSRSAGSRYSKKGSRGRWYWKKGCSRATSKKGRMMW 189

Qy        606 CGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGGA 665
              :::|::::  : : : :      :::::  : :  ::    : :::::||::   |::
Db        190 MKKGSRRRATSRYGMMSSMYGASKRMSSMCSASTRMSSASCMMYMMMSAGSYASCAWKMS 249

Qy        666 ACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGG 699
              :::|   :||: :: :| : :::  |::: |::
Db        250 KYRRCAKWSCTYSWYMRASMKKSKYCAWSRKGSK 283


RESULT 14
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350003
;  CURRENT APPLICATION NUMBER: US/09/007,005B
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

  Query Match             1.5%;  Score 40.2;  DB 3;  Length 289;
  Best Local Similarity   5.6%;  Pred. No. 0.079;
  Matches   12;  Conservative   97;  Mismatches  106;  Indels    0;  Gaps    0;

Qy        808 AAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTG 867
              :|: :|:   :::   ::  : ::: : ::    ::: : :::    ::   :::    :
Db         51 RARURGRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS 110

Qy        868 GGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAAC 927
              :   :: : ::   :::    :: : :::    :   : : : ::: :  :: : ::
Db        111 RNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS 170

Qy        928 GATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAG 987
              :  ::: :   : :  :  :   : : : :  ::: : :::   ::  : :|: : :|

Db        171 RNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRCRARGRCRURG 230

Qy        988 GCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGA 1022
              :|:| : |: :|::  : :  |: : :| :  | |
Db        231 RCRGRURARARCRURCRURURGRCRGRCRUAAAAA 265


RESULT 15
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350007
;  CURRENT APPLICATION NUMBER: US/09/244,796
;  CURRENT FILING DATE: 1999-02-05
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  EARLIER APPLICATION NUMBER: 09/007,005
;  EARLIER FILING DATE: 1998-01-14
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

  Query Match             1.5%;  Score 40.2;  DB 3;  Length 289;
  Best Local Similarity   5.6%;  Pred. No. 0.079;
  Matches   12;  Conservative   97;  Mismatches  106;  Indels    0;  Gaps    0;

Qy        808 AAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTG 867
              :|: :|:   :::   ::  : ::: : ::    ::: : :::    ::   :::    :
Db         51 RARURGRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS 110

Qy        868 GGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAAC 927
              :   :: : ::   :::    :: : :::    :   : : : ::: :  :: : ::
Db        111 RNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS 170

Qy        928 GATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAG 987
              :  ::: :   : :  :  :   : : : :  ::: : :::   ::  : :|: : :|

Db        171 RNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRCRARGRCRURG 230

Qy        988 GCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGA 1022
              :|:| : |: :|::  : :  |: : :| :  | |
Db        231 RCRGRURARARCRURCRURURGRCRGRCRUAAAAA 265


Search completed: April 23, 2004, 18:09:01
Job time : 191.89 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 23, 2004, 14:17:01 ; Search time 1241.38 Seconds
(without alignments)
9479.366 Million cell updates/sec

Title: US-09-787-016A-1
Perfect score: 2610
Sequence: 1 ctcggtggccgtccgcccac.....actcttaagatcatatcctg 2610

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2907579 seqs, 2254313464 residues

Total number of hits satisfying chosen parameters: 5815158

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1564.4	59.9	1764	9	US-09-925-301-327	Sequence 327, App
c 2	645	24.7	645	15	US-10-029-386-25091	Sequence 25091, A
3	435	16.7	478	10	US-09-918-995-19127	Sequence 19127, A
c 4	152.6	5.8	536	15	US-10-029-386-11381	Sequence 11381, A
c 5	91.2	3.5	226	16	US-10-131-827-8445	Sequence 8445, Ap
6	65	2.5	5975	10	US-09-873-367C-112	Sequence 112, App
7	65	2.5	6256	15	US-10-037-270-269	Sequence 269, App
8	65	2.5	6256	16	US-10-117-722-269	Sequence 269, App
9	65	2.5	6948	10	US-09-930-213-315	Sequence 315, App
10	60	2.3	60	10	US-09-908-975-6269	Sequence 6269, Ap
11	52.8	2.0	469	13	US-10-085-783A-56748	Sequence 56748, A
12	52.8	2.0	469	16	US-10-242-535A-56748	Sequence 56748, A
13	52.8	2.0	483	9	US-09-864-761-2140	Sequence 2140, Ap
14	51.4	2.0	915	9	US-09-764-864-471	Sequence 471, App
15	51.4	2.0	2320	15	US-10-175-523-142	Sequence 142, App
16	50	1.9	50	16	US-10-131-827-916	Sequence 916, App
17	48	1.8	594	13	US-10-142-426-10	Sequence 10, Appl
18	48	1.8	594	15	US-10-123-155-10	Sequence 10, Appl
19	48	1.8	594	15	US-10-146-731-10	Sequence 10, Appl
20	48	1.8	594	15	US-10-140-472-10	Sequence 10, Appl
21	48	1.8	594	15	US-10-141-761-10	Sequence 10, Appl
22	48	1.8	594	15	US-10-142-885-10	Sequence 10, Appl
23	48	1.8	594	15	US-10-158-790-10	Sequence 10, Appl
24	48	1.8	594	16	US-10-137-871-10	Sequence 10, Appl
25	48	1.8	594	16	US-10-140-923-10	Sequence 10, Appl
26	48	1.8	594	16	US-10-141-756-10	Sequence 10, Appl
27	48	1.8	594	16	US-10-141-759-10	Sequence 10, Appl
28	48	1.8	594	16	US-10-140-805-10	Sequence 10, Appl
29	48	1.8	594	16	US-10-140-864-10	Sequence 10, Appl
30	44.4	1.7	7386	10	US-09-819-104A-6	Sequence 6, Appli
31	44.4	1.7	7534	13	US-10-087-192-650	Sequence 650, App
32	44.4	1.7	8544	10	US-09-819-104A-4	Sequence 4, Appli
33	44.4	1.7	91141	13	US-10-087-192-649	Sequence 649, App
c 34	42.6	1.6	16442	16	US-10-374-077-208	Sequence 208, App
35	42	1.6	2085	17	US-10-311-623-18	Sequence 18, Appl
36	42	1.6	2771	13	US-10-112-944-494	Sequence 494, App
37	42	1.6	3195	15	US-10-205-823-143	Sequence 143, App
38	42	1.6	3195	15	US-10-285-976-54	Sequence 54, Appl
39	42	1.6	3195	16	US-10-295-027-1127	Sequence 1127, Ap
40	42	1.6	3206	13	US-10-112-944-6	Sequence 6, Appli
c 41	42	1.6	1223197	13	US-10-027-632-179264	Sequence 179264,
c 42	42	1.6	1223197	16	US-10-027-632-179264	Sequence 179264,
c 43	41.8	1.6	636	9	US-09-770-149-640	Sequence 640, App
c 44	41.8	1.6	659158	9	US-09-771-208-20	Sequence 20, Appl
c 45	41.2	1.6	4056	15	US-10-156-761-2317	Sequence 2317, Ap

ALIGNMENTS

RESULT 1
US-09-925-301-327
; Sequence 327, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 327
; LENGTH: 1764
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1398)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1758)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1759)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1762)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-327

Query Match 59.9%; Score 1564.4; DB 9; Length 1764;
Best Local Similarity 99.6%; Pred. No. 0;

```
  Matches 1610;  Conservative    0;  Mismatches    2;  Indels    5;  Gaps    4;

Qy        996 GGACATCAAAGATGAGGAGCCTGGAGAC-TTGGGCCGACCGAAGCCTGAATGTGAGGGTT 1054
              ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db          1 GGACATCAAAGATGAGGAGCCTGGAGACTTTGGGCCGACCGAAGCCTGAATGTGAGGGTT 60

Qy       1055 ACGACCCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTT 1114
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACGACCCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTT 120

Qy       1115 GCTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAG 1174
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAG 180

Qy       1175 GGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACTGCACCATTCTGCAAG 1234
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACTGCACCATTCTGCAAG 240

Qy       1235 TGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAG 1294
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAG 300

Qy       1295 ATGCTGATGGCACCGATTGTACAAGTATAGGAACAATAGAGCAGAAGTCTAGCGAAGACC 1354
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ATGCTGATGGCACCGATTGTACAAGTATAGGAACAATAGAGCAGAAGTCTAGCGAAGACC 360

Qy       1355 AAGGGATAAAGGGTAGAATTGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGA 1414
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAGGGATAAAGGGTAGAATTGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGA 420

Qy       1415 TCTTCCAGCCTGTGATAGAGGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTC 1474
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCTTCCAGCCTGTGATAGAGGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTC 480

Qy       1475 ACGTGGCGCAGCCCGACTCGGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAG 1534
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ACGTGGCGCA-CCCGACTCGGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAG 539

Qy       1535 CGACAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGA 1594
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 CGACAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGA 599

Qy       1595 AGATGAAGCCAGAGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCT 1654
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 AGATGAAGCCAGAGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCT 659

Qy       1655 CTTCTGTGCACAAGAGACCAGCTCCAGAAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGG 1714
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 CTTCTGTGCACAAGAGACCAGCTCCAGAAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGG 719

Qy       1715 TGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGT 1774
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 TGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGT 779

Qy       1775 CGTGGGCGAGCGATCACAATTACAATGCAGTAAAGCCAGAAAAGACTGCTGCTCCCTCGC 1834
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        780 CGTGGGCGAGCGATCACAATTACAATGCAGTAAAGCCAGAAAAGACTGCTGCTCCCTCGC 839

Qy       1835 CGTCACTGTTGTATAAATGTATGTATCACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTT 1894
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 CGTCACTGTTGTATAAATGTATGTATCACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTT 899

Qy       1895 CTTTCTGGATAGCCATCCCCTGGGCCTGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAG 1954
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        900 CTTTCTGGATAGCCATCCCCTGGGCCTGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAG 959

Qy       1955 CTGATCACAGACACCGGCTGCACCATCAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTC 2014
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 CTGATCACAGACACCGGCTGCACCATCAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTC 1019

Qy       2015 CTGCTGCCCTGTGTCCATCCCTAGTCTGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTG 2074
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1020 CTGCTGCCCTGTGTCCATCCCTAGTCTGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTG 1079

Qy       2075 TAAACCTCACTGGTGAACGTTCACCTTAATGATTGATTCTTTAATCTCTGTTTTCACTCT 2134
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 TAAACCTCACTGGTGAACGTTCACCTTAATGATTGATTCTTTAATCTCTGTTTTCACTCT 1139

Qy       2135 CAGGCTCTGGTAAGTATTTGTATTCTCTTCATCCCAGTCTGATTGCATAGCCACACTGCC 2194
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       1140 CAGGCTCTGGTAAGTATTCGTATTCTCTTCATCCCAGTCTGATTGCATAGCCACACTGCC 1199

Qy       2195 CGGCACGCCACATCCACCCCTGTCTGCACATGAGTTGTTCTGACAACAGCGCTGTATACG 2254
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1200 CGGCACGCCACATCCACCCCTGTCTGCACATGAGTTGTTCTGACAACAGCGCTGTATACG 1259

Qy       2255 CTTCAGTTTTTCCACATTGTCCACGGCCAGCACATGAAAGCATCACTTCTTTTTTATGTT 2314
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1260 CTTCAGTTTTTCCACATTGTCCACGGCCAGCACATGAAAGCATCACTTCTTTTTTATGTT 1319

Qy       2315 GTGGGAATCTTTGCAAGTTAGTGTTGCATCTGATTTTCAGGTGTACATTTATTTTTGACT 2374
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1320 GTGGGAATCTTTGCAAGTTAGTGTTGCATCTGATTTTCAGGTGTACATTTATTTTTGACT 1379

Qy       2375 GGGCAGATAGGGGATTTTTTTTTTTTTCCATGTCCGATTCACACGCTACACACCCACATG 2434
              ||||||||||||| |||| ||||||||||||||||||||||||||||||||||||||||||
Db       1380 GGGCAGATAGGGGA--TTTTNTTTTTTCCATGTCCGATTCACACGCTACACACCCACATG 1437

Qy       2435 AACACATTCGAACTTCGAAGG-CACACACTCCTGCTTCATAGGCCCCACGGTAAGTGAGT 2493
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db       1438 AACACATTCGAACTTCGAAGGCCACACACTCCTGCTTCATAGGCCCCACGGTAAGTGAGT 1497

Qy       2494 TCACACCTAGAACACTGTCCTGACCGCAGGACGCGTGCCTTGGACTTGGTATTCTACATG 2553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1498 TCACACCTAGAACACTGTCCTGACCGCAGGACGCGTGCCTTGGACTTGGTATTCTACATG 1557

Qy       2554 TGACTGGCTTTCTTGCCCTCGTCTCTTGAATGTTTAGACTCTTAAGATCATATCCTG 2610
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1558 TGACTGGCTTTCTTGCCCTCGTCTCTTGAATGTTTAGACTCTTAAGATCATATCCTG 1614
```

RESULT 2
US-10-029-386-25091/c
; Sequence 25091, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GI
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 25091
; LENGTH: 645
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL117379.14
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: P23583, EVALUE 2.50e-01
; OTHER INFORMATION: EST_HUMAN HIT: AL580669.1, EVALUE 0.00e+00
; OTHER INFORMATION: NT HIT: gi16193858, EVALUE 0.00e+00
US-10-029-386-25091

```
  Query Match           24.7%;  Score 645;  DB 15;  Length 645;
  Best Local Similarity  100.0%;  Pred. No. 1.5e-196;
  Matches  645;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        263 GTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTAAGGCCATCAAACCCACCA 322
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        645 GTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTAAGGCCATCAAACCCACCA 586
```

```
Qy        323 GCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTATCGCCAAGCGAGAGGGCG 382
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        585 GCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTATCGCCAAGCGAGAGGGCG 526

Qy        383 CAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCACAGCAGCAGCTGGGCCTGT 442
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        525 CAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCACAGCAGCAGCTGGGCCTGT 466

Qy        443 CCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGACCA 502
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        465 CCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGACCA 406

Qy        503 TTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGCCCA 562
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        405 TTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGCCCA 346

Qy        563 CGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCAGCGTGGAAAGCGCTTCTG 622
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        345 CGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCAGCGTGGAAAGCGCTTCTG 286

Qy        623 AGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGGAACGACCAGCCTCTTCTG 682
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        285 AGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGGAACGACCAGCCTCTTCTG 226

Qy        683 AAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGGCCTGA 742
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        225 AAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGGCCTGA 166

Qy        743 CCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGAGGC 802
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        165 CCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGAGGC 106

Qy        803 CCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCGAGA 862
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        105 CCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCGAGA 46

Qy        863 CTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGCCCA 907
              |||||||||||||||||||||||||||||||||||||||||||||
Db         45 CTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGCCCA 1
```

RESULT 3
US-09-918-995-19127
; Sequence 19127, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 19127
; LENGTH: 478
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(478)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-19127

Query Match 16.7%; Score 435; DB 10; Length 478;
Best Local Similarity 100.0%; Pred. No. 6.8e-129;
Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        905 CCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACA 964
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         44 CCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACA 103

Qy        965 GAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACT 1024
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        104 GAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACT 163

Qy       1025 TGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCC 1084
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        164 TGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCC 223

Qy       1085 GCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATG 1144
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        224 GCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATG 283

Qy       1145 GCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACT 1204
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        284 GCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACT 343

Qy       1205 ATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAG 1264
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        344 ATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAG 403

Qy       1265 ATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAG 1324
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        404 ATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAG 463

Qy       1325 GAACAATAGAGCAGA 1339
              |||||||||||||||
Db        464 GAACAATAGAGCAGA 478
```

RESULT 4
US-10-029-386-11381/c
; Sequence 11381, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GI
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 11381
; LENGTH: 536
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL117379.14
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: P80398, EVALUE 3.80e+00
; OTHER INFORMATION: NT HIT: gi14769835, EVALUE 1.00e-78
; OTHER INFORMATION: EST_HUMAN HIT: BF205331.1, EVALUE 2.00e-78
US-10-029-386-11381

Query Match 5.8%; Score 152.6; DB 15; Length 536;
Best Local Similarity 94.6%; Pred. No. 9.5e-38;
Matches 158; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```
Qy        244 GCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCT 303
              |||| || | |       ||||||||||||||||||||||||||||||||||||||||||||
Db        167 GCTTACACTTTACCGCTAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCT 108

Qy        304 AAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        107 AAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACT 48

Qy        364 ATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGA 410
              |||||||||||||||||||||||||||||||||||||||||||||
Db         47 ATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGA 1
```

```
RESULT 5
US-10-131-827-8445/c
; Sequence 8445, Application US/10131827
; Publication No. US20040009479A1
; GENERAL INFORMATION:
;  APPLICANT: Wohlgemuth, Jay
;  APPLICANT:  Fry, Kirk
;  APPLICANT:  Woodward, Robert
;  APPLICANT:  Ly, Ngoc
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING AND MONITORING AUTOIMMUNE
;  TITLE OF INVENTION:  CHRONIC INFLAMMATORY DISEASES
;  FILE REFERENCE: 506612000120
;  CURRENT APPLICATION NUMBER: US/10/131,827
;  CURRENT FILING DATE:  2002-09-06
;  PRIOR APPLICATION NUMBER: US 10/006,290
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: US 60/296,764
;  PRIOR FILING DATE: 2001-06-08
;  NUMBER OF SEQ ID NOS: 9090
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8445
;   LENGTH: 226
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-131-827-8445

  Query Match             3.5%;  Score 91.2;  DB 16;  Length 226;
  Best Local Similarity   72.5%;  Pred. No. 3.4e-18;
  Matches  145;  Conservative    0;  Mismatches   53;  Indels    2;  Gaps    2;

Qy       2153 TGTATTCTCTTCATCCCAGTCTGATTGCATAGCCACACTGCCCGGCACGCCACATCCACC 2212
             | |||| | |||||||| ||  | || || |  | || ||| | |  | | ||||||||
Db        217 TTTATTTTTTTCATCCCTGTGAGTTTTCAGATTCTCATTGCTCCGTCCTCAACATCCACT 158

Qy       2213 CCTGTCTGCACATGAGTTGTTCTGACAACAGCGCTGTATACGCTTCAG-TTTTTCCACAT 2271
             || || |||||||||  ||||||     || | |||||| |  ||||| ||||| | |||
Db        157 CCAGTTTGCACATGACATGTTCTACATGCATCACTGTATTCATTTCAGTTTTTTTCTCAT 98

Qy       2272 TGTCCACGGCCAGCACATGAAAGCAT-CACTTCTTTTTTATGTTGTGGGAATCTTTGCAA 2330
             |||||| ||||| |||| || | || ||  | ||||| |||||||| |||||| ||||
Db         97 AGTCCACTGCCAGAACATTAATGTATACATATGTTTTTAATGTTGTGGTAATCTTGGCAA 38

Qy       2331 GTTAGTGTTGCATCTGATTT 2350
             |||||| | ||||  ||||
Db         37 GTTAGTATAGCATACAATTT 18


RESULT 6
US-09-873-367C-112
; Sequence 112, Application US/09873367C
; Publication No. US20030165839A1
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  APPLICANT:  Soppet, Daniel
;  APPLICANT:  Endress, Gregory
;  APPLICANT:  Augustus, Meena
;  APPLICANT:  Ebner, Reinhard
;  APPLICANT:  Carter, Kenneth
;  TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
;  TITLE OF INVENTION:  Signature Gene Sets
;  FILE REFERENCE: 689290-64
;  CURRENT APPLICATION NUMBER: US/09/873,367C
;  CURRENT FILING DATE:  2003-04-29
;  PRIOR APPLICATION NUMBER: U.S. 60/236,891
;  PRIOR FILING DATE: 2000-09-29
;  PRIOR APPLICATION NUMBER: U.S. 60/236,842
;  PRIOR FILING DATE: 2000-09-29
;  PRIOR APPLICATION NUMBER: U.S. 60/244,867
;  PRIOR FILING DATE: 2000-11-01
;  PRIOR APPLICATION NUMBER: U.S. 60/245,084
;  PRIOR FILING DATE: 2000-11-01
;  NUMBER OF SEQ ID NOS: 1067
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 112
;   LENGTH: 5975
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-873-367C-112

  Query Match             2.5%;  Score 65;  DB 10;  Length 5975;
  Best Local Similarity   62.7%;  Pred. No. 9.1e-09;
  Matches  101;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy       1060 CCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGT 1119
             |||| |   | || | |  |||||    | || ||   |||||||||||||| || |||||
Db       1200 CCCAGCAAGCAGTGTGGGTTTTGCAAAAAACCACATGGCAACAGGTTTATGGTTGGCTGT 1259

Qy       1120 GACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGG 1179
             |   | |||||| || |||||||||||| |||||||| || |   |    ||| ||    |
Db       1260 GGGAGATGTGATGACTGGTTTCATGGTGATTGTGTTGGGTTAAGTCTTTCTCAAGCACAG 1319

Qy       1180 CTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACTG 1220
             |   |||    | | |   |||| ||| ||||   ||| ||
Db       1320 CAGATGGGCGAGGAAGACAAAGAATATGTCTGTGTAAAATG 1360


RESULT 7
US-10-037-270-269
; Sequence 269, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Ren, Peiyan
;  APPLICANT:  Chen, Rui-hong
;  APPLICANT:  Zhao, Qing A.
;  APPLICANT:  Wehrman, Tom
;  APPLICANT:  Xue, Aidong J.
;  APPLICANT:  Yang, Yonghong
;  APPLICANT:  Wang, Jian-Rui
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Ma, Yunqing
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Wang, Zhiwei
;  APPLICANT:  Tillinghast, John
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 784CIP2B
;  CURRENT APPLICATION NUMBER: US/10/037,270
;  CURRENT FILING DATE:  2002-01-04
;  PRIOR APPLICATION NUMBER: 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 1104
;  SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 269
;   LENGTH: 6256
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (257)..(5443)
US-10-037-270-269

  Query Match             2.5%;  Score 65;  DB 15;  Length 6256;
  Best Local Similarity   62.7%;  Pred. No. 9.4e-09;
  Matches  101;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy       1060 CCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGT 1119
             |||| |   | || | |  |||||    | || ||   |||||||||||||| || |||||
Db       1469 CCCAGCAAGCAGTGTGGGTTTTGCAAAAAACCACATGGCAACAGGTTTATGGTTGGCTGT 1528
```

```
Qy       1120 GACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGG 1179
             |   | |||||  || |||||||||| ||||||||  ||   |    |    ||| ||    |
Db       1529 GGGAGATGTGATGACTGGTTTCATGGTGATTGTGTTGGGTTAAGTCTTTCTCAAGCACAG 1588

Qy       1180 CTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACTG 1220
             |   |||    | | |   |||| ||| ||||   ||| ||
Db       1589 CAGATGGGCGAGGAAGACAAAGAATATGTCTGTGTAAAATG 1629


RESULT 8
US-10-117-722-269
; Sequence 269, Application US/10117722
; Publication No. US20030219744A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. US20030219744A1el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 784CIP2BCIP
;  CURRENT APPLICATION NUMBER: US/10/117,722
;  CURRENT FILING DATE:  2002-04-04
;  PRIOR APPLICATION NUMBER: 09/620,312
;  PRIOR FILING DATE: 2000-07-19
;  PRIOR APPLICATION NUMBER: 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 1104
;  SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 269
;   LENGTH: 6256
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (257)..(5443)
US-10-117-722-269

  Query Match             2.5%;  Score 65;  DB 16;  Length 6256;
  Best Local Similarity   62.7%;  Pred. No. 9.4e-09;
  Matches  101;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy       1060 CCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGT 1119
             |||| |   | || | |  |||||     | || ||   |||||||||||||| || |||||
Db       1469 CCCAGCAAGCAGTGTGGGTTTTGCAAAAAACCACATGGCAACAGGTTTATGGTTGGCTGT 1528

Qy       1120 GACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGG 1179
             |   | |||||  || |||||||||| ||||||||  ||   |    |    ||| ||    |
Db       1529 GGGAGATGTGATGACTGGTTTCATGGTGATTGTGTTGGGTTAAGTCTTTCTCAAGCACAG 1588

Qy       1180 CTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACTG 1220
             |   |||    | | |   |||| ||| ||||   ||| ||
Db       1589 CAGATGGGCGAGGAAGACAAAGAATATGTCTGTGTAAAATG 1629


RESULT 9
US-09-930-213-315
; Sequence 315, Application US/09930213
; Publication No. US20030170625A1
; GENERAL INFORMATION:
;  APPLICANT: ROSENTHAL, ANDRE
;  APPLICANT:  HINZMANN, BERND
;  APPLICANT:  SCHAFER, REINHARD
;  APPLICANT:  ZUBER, JOHANNES
;  APPLICANT:  TCHE-NITSE, OLEG
;  APPLICANT:  GRIPS, MARTIN
;  APPLICANT:  HELLNEGEL, MARTIN
;  APPLICANT:  SCHMITZ, ANNE-CHANTAL
;  APPLICANT:  SERS, CHRISTINE
;  TITLE OF INVENTION: DETECTION OF DIFFERENTIAL GENE EXPRESSIONS
;  FILE REFERENCE: ALBRE-14
;  CURRENT APPLICATION NUMBER: US/09/930,213
;  CURRENT FILING DATE:  2001-01-31
;  PRIOR APPLICATION NUMBER: DE 10004102.7
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 885
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 315
;   LENGTH: 6948
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-930-213-315

  Query Match             2.5%;  Score 65;  DB 10;  Length 6948;
  Best Local Similarity   62.7%;  Pred. No. 1e-08;
  Matches  101;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy       1060 CCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGT 1119
             |||| |   | || | |  |||||     | || ||   |||||||||||||| || |||||
Db       2173 CCCAGCAAGCAGTGTGGGTTTTGCAAAAAACCACATGGCAACAGGTTTATGGTTGGCTGT 2232

Qy       1120 GACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGG 1179
             |   | |||||  || |||||||||| ||||||||  ||   |    |    ||| ||    |
Db       2233 GGGAGATGTGATGACTGGTTTCATGGTGATTGTGTTGGGTTAAGTCTTTCTCAAGCACAG 2292

Qy       1180 CTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACTG 1220
             |   |||    | | |   |||| ||| ||||   ||| ||
Db       2293 CAGATGGGCGAGGAAGACAAAGAATATGTCTGTGTAAAATG 2333


RESULT 10
US-09-908-975-6269
; Sequence 6269, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
;  APPLICANT: SHOSHAN, Avi
;  APPLICANT:  WASSERMAN, Alon
;  APPLICANT:  MINTZ, Eli
;  APPLICANT:  MINTZ, Liat
;  APPLICANT:  FAIGLER, Simchon
;  TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE \
;  TITLE OF INVENTION:  THAT POPULATE A TRANSCRIPTOME
;  FILE REFERENCE: 36688-0005
;  CURRENT APPLICATION NUMBER: US/09/908,975
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: US 60/287,724
;  PRIOR FILING DATE: 2001-05-02
;  PRIOR APPLICATION NUMBER: US 60/221,607
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 32337
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6269
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-908-975-6269

  Query Match             2.3%;  Score 60;  DB 10;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-08;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1220 GCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTA 1279
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTA 60


RESULT 11
US-10-085-783A-56748
; Sequence 56748, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
;  APPLICANT: ChondroGene Inc.
```

```
;  APPLICANT:  Liew, C.C.
;  TITLE OF INVENTION: Compositions and Methods Relatiing to Osteoarthritis
;  FILE REFERENCE: 4231/2002
;  CURRENT APPLICATION NUMBER: US/10/085,783A
;  CURRENT FILING DATE:  2002-02-28
;  PRIOR APPLICATION NUMBER: US 60/305,340
;  PRIOR FILING DATE: 2001-07-13
;  PRIOR APPLICATION NUMBER: US 60/275,017
;  PRIOR FILING DATE: 2001-03-12
;  PRIOR APPLICATION NUMBER: US 60/271,955
;  PRIOR FILING DATE: 2001-02-28
;  NUMBER OF SEQ ID NOS: 58994
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 56748
;   LENGTH: 469
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (5)..(5)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (30)..(30)
;   OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-56748

  Query Match             2.0%;  Score 52.8;  DB 13;  Length 469;
  Best Local Similarity   65.0%;  Pred. No. 1.4e-05;
  Matches   78;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy       1101 CAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCAT 1160
             |||||||||| || ||||||   | ||||| || |||||||||| ||||||||| || |
Db        116 CAGGTTTATGGTTGGCTGTGGGAGATGTGATGACTGGTTTCATGGTGATTGTGTTGGGTT 175

Qy       1161 TTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACTG 1220
                 |  |||  ||    ||  |||    | | |   |||| ||| ||||   ||| ||
Db        176 AAGTCTTTCTCAAGCACAGCAGATGGGCGAGGAAGACAAAGAATATGTCTGTGTAAAATG 235


RESULT 12
US-10-242-535A-56748
; Sequence 56748, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
;  APPLICANT: ChondroGene Inc.
;  APPLICANT:  Liew, C.C.
;  TITLE OF INVENTION: Compositions and Methods Relatiing to Osteoarthritis
;  FILE REFERENCE: 4231/2005
;  CURRENT APPLICATION NUMBER: US/10/242,535A
;  CURRENT FILING DATE:  2002-09-12
;  PRIOR APPLICATION NUMBER: US 10/085,783
;  PRIOR FILING DATE: 2002-02-28
;  PRIOR APPLICATION NUMBER: US 60/305,340
;  PRIOR FILING DATE: 2001-07-13
;  PRIOR APPLICATION NUMBER: US 60/275,017
;  PRIOR FILING DATE: 2001-03-12
;  PRIOR APPLICATION NUMBER: US 60/271,955
;  PRIOR FILING DATE: 2001-02-28
;  NUMBER OF SEQ ID NOS: 58994
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 56748
;   LENGTH: 469
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (5)..(5)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (30)..(30)
;   OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-56748

  Query Match             2.0%;  Score 52.8;  DB 16;  Length 469;
  Best Local Similarity   65.0%;  Pred. No. 1.4e-05;
  Matches   78;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy       1101 CAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCAT 1160
             |||||||||| || ||||||   | ||||| || |||||||||| ||||||||| || |
Db        116 CAGGTTTATGGTTGGCTGTGGGAGATGTGATGACTGGTTTCATGGTGATTGTGTTGGGTT 175

Qy       1161 TTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACTG 1220
                 |  |||  ||    ||  |||    | | |   |||| ||| ||||   ||| ||
Db        176 AAGTCTTTCTCAAGCACAGCAGATGGGCGAGGAAGACAAAGAATATGTCTGTGTAAAATG 235


RESULT 13
US-09-864-761-2140
; Sequence 2140, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 2140
;   LENGTH: 483
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL050329.9
```

```
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL =   1
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.2
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =   1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.94
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.9
US-09-864-761-2140

  Query Match             2.0%;  Score 52.8;  DB 9;  Length 483;
  Best Local Similarity   65.0%;  Pred. No. 1.4e-05;
  Matches   78;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy       1101 CAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCAT 1160
              |||||||||| || ||||||   | ||||| || ||||||||||| |||||||| ||   |
Db        219 CAGGTTTATGGTTGGCTGTGGGAGATGTGATGACTGGTTTCATGGTGATTGTGTTGGGTT 278

Qy       1161 TTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACTG 1220
                |   ||| ||   ||  |||   | | |  |||| ||| ||||  ||| ||
Db        279 AAGTCTTTCTCAAGCACAGCAGATGGGCGAGGAAGACAAAGAATATGTCTGTGTAAAATG 338


RESULT 14
US-09-764-864-471
; Sequence 471, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 471
;   LENGTH: 915
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (894)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (903)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (905)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (914)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-471

  Query Match             2.0%;  Score 51.4;  DB 9;  Length 915;
  Best Local Similarity   69.3%;  Pred. No. 6.1e-05;
  Matches   70;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       1067 CCCTGTATTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCT 1126
              || | || ||||| |||||| | ||  ||| |||| | || |||||  | ||||||  ||
Db        305 CCATCTACTGCATCTGCCGCAAACCGGACATCAACTGCTTCATGATCGGGTGTGACAACT 364

Qy       1127 GTGAAGAATGGTTTCATGGCGATTGTGTGGGCATTTCTGAG 1167
              |  | || ||||| ||||| || ||  |  | ||  |||||
Db        365 GCAATGAGTGGTTCCATGGGGACTGCATCCGGATCACTGAG 405


RESULT 15
US-10-175-523-142
; Sequence 142, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
;  APPLICANT: Brockman, Jeffrey
;  APPLICANT:  Evans, David
;  APPLICANT:  Hook, Derek
;  APPLICANT:  Klimczak, Leszek
;  APPLICANT:  Laeng, Pascal
;  APPLICANT:  Palfreyman, Michael
;  APPLICANT:  Rajan, Prithi
;  TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
;  FILE REFERENCE: 3235/1J795-US3
;  CURRENT APPLICATION NUMBER: US/10/175,523
;  CURRENT FILING DATE:  2002-06-18
;  PRIOR APPLICATION NUMBER: US 60/299,151
;  PRIOR FILING DATE: 2001-06-18
;  PRIOR APPLICATION NUMBER: US 60/317,828
;  PRIOR FILING DATE: 2001-09-07
;  PRIOR APPLICATION NUMBER: US 60/325,150
;  PRIOR FILING DATE: 2001-09-25
;  PRIOR APPLICATION NUMBER: US 60/333,047
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: US 60/349,936
;  PRIOR FILING DATE: 2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/361,834
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 197
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 142
;   LENGTH: 2320
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-175-523-142

  Query Match             2.0%;  Score 51.4;  DB 15;  Length 2320;
  Best Local Similarity   69.3%;  Pred. No. 0.00012;
  Matches   70;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       1067 CCCTGTATTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCT 1126
              || | || ||||| |||||| | ||  ||| |||| | || |||||  | ||||||  ||
Db        133 CCATCTACTGCATCTGCCGCAAACCGGACATCAACTGCTTCATGATCGGGTGTGACAACT 192

Qy       1127 GTGAAGAATGGTTTCATGGCGATTGTGTGGGCATTTCTGAG 1167
              |  | || ||||| ||||| || ||  |  | ||  |||||
Db        193 GCAATGAGTGGTTCCATGGGGACTGCATCCGGATCACTGAG 233


Search completed: April 23, 2004, 23:50:37
Job time : 1250.38 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 23, 2004, 07:20:53 ; Search time 6433.74 Seconds
(without alignments)
12114.300 Million cell updates/sec

Title: US-09-787-016A-1
Perfect score: 2610
Sequence: 1 ctcggtggccgtccgcccac.....actcttaagatcatatcctg 2610

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	Query Match %	Length	DB	ID	Description
	1	2539	97.3	2752	11	BC014615	BC014615 Homo sapi
	2	1025.2	39.3	4619	11	AK042474	AK042474 Mus muscu
c	3	1007.2	38.6	1086	9	AL578222	AL578222 AL578222
	4	1002.6	38.4	4651	11	BC060249	BC060249 Mus muscu
	5	957.2	36.7	1093	9	AL555810	AL555810 AL555810
	6	910.6	34.9	1033	9	AL580669	AL580669 AL580669
c	7	905.6	34.7	1201	9	AL576399	AL576399 AL576399
	8	900.2	34.5	1010	13	BX380892	BX380892 BX380892
	9	868	33.3	1201	9	AL514138	AL514138 AL514138
	10	863.6	33.1	1201	9	AL553387	AL553387 AL553387
c	11	845.2	32.4	1018	9	AL550675	AL550675 AL550675
	12	838.2	32.1	1050	13	BX404270	BX404270 BX404270
c	13	830.4	31.8	850	14	CD364822	CD364822 UI-H-FT2-
	14	817.8	31.3	1614	11	AK014422	AK014422 Mus muscu
	15	813.4	31.2	1119	9	AL550431	AL550431 AL550431
	16	808.6	31.0	1640	11	BC029110	BC029110 Mus muscu
	17	798.6	30.6	898	13	BQ646878	BQ646878 AGENCOURT
c	18	792.8	30.4	939	9	AL571596	AL571596 AL571596
c	19	786.6	30.1	891	13	BX350413	BX350413 BX350413
c	20	786.4	30.1	1040	9	AL580668	AL580668 AL580668
	21	786.2	30.1	925	9	AL546121	AL546121 AL546121
	22	765.8	29.3	1201	13	BX442517	BX442517 BX442517
	23	753.4	28.9	781	13	BQ921479	BQ921479 AGENCOURT
c	24	740.8	28.4	1061	9	AL574291	AL574291 AL574291
	25	738.8	28.3	842	13	BQ437271	BQ437271 AGENCOURT
	26	733.2	28.1	866	13	BQ212019	BQ212019 AGENCOURT
c	27	722.4	27.7	1201	9	AL514137	AL514137 AL514137
	28	717	27.5	1045	9	AL550704	AL550704 AL550704
c	29	708.4	27.1	740	13	BU689908	BU689908 UI-CP-FN0
	30	693.8	26.6	779	12	BG825278	BG825278 602747725
	31	673.2	25.8	842	12	BG757160	BG757160 602710529
	32	670.6	25.7	839	12	BG828423	BG828423 602751590
	33	668.6	25.6	1097	12	BG290255	BG290255 602384956
c	34	668.2	25.6	900	13	BX413695	BX413695 BX413695
	35	648.6	24.9	760	10	BE298017	BE298017 601118492
	36	645	24.7	740	12	BI913500	BI913500 603179127
	37	634.4	24.3	963	13	BU526593	BU526593 AGENCOURT
	38	629.6	24.1	772	10	BE299127	BE299127 601118879
	39	626.8	24.0	692	10	BE393471	BE393471 601310142
	40	625	23.9	635	13	BQ639998	BQ639998 he22h10.y
	41	615.6	23.6	653	10	AW373892	AW373892 QV0-BT053
	42	615.6	23.6	653	10	AW373929	AW373929 QV3-BT053
	43	614.4	23.5	923	10	BF308603	BF308603 601890469
	44	612.4	23.5	1036	13	BQ881380	BQ881380 AGENCOURT
	45	608.8	23.3	868	10	BF306721	BF306721 601888796

ALIGNMENTS

RESULT 1
BC014615
LOCUS BC014615 2752 bp mRNA linear HTC 17-DEC-2003
DEFINITION Homo sapiens cDNA clone IMAGE:3532298, containing frame-shift errors.
ACCESSION BC014615
VERSION BC014615.2 GI:33869935
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2752)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,

Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED 12477932
REFERENCE 2 (bases 1 to 2752)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (24-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT On Aug 19, 2003 this sequence version replaced gi:15779097.
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbiology.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 11 Row: j Column: 11
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 18375616
This clone has the following problem: frame shifted.

FEATURES Location/Qualifiers
source 1. .2752
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3532298"
/tissue_type="Muscle, rhabdomyosarcoma"
/clone_lib="NIH_MGC_17"
/lab_host="DH10B-R"
/note="Vector: pOTB7"

ORIGIN

Query Match 97.3%; Score 2539; DB 11; Length 2752;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2573; Conservative 0; Mismatches 0; Indels 4; Gaps 3;

```
Qy      36 GAAATGGCTGCGAGACCCTAGAGGCCTGCGGAGCTTACTCCACGGGAACAGCCTCTAGAT 95
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GAAATGGCTGCGAGACCCTAGAGGCCTGCGGAGCTTACTCCACGGGAACAGCCTCTAGAT 60

Qy      96 AATCTGAGTTGTTGAAAATACGAAGCCTGTTACTCGTGAACAGTGGCTGACAACAGTGTT 155
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AATCTGAGTTGTTGAAAATACGAAGCCTGTTACTCGTGAACAGTGGCTGACAACAGTGTT 120

Qy     156 GTTGTGAGCCTGGCTGTCTGCTTGGACCCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGT 215
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GTTGTGAGCCTGGCTGTCTGCTTGGACCCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGT 180

Qy     216 ATTTAGGATTTCAGGGAAAAGTGTCCAAGCTTTCAGTGTTGGAGCAGGTATGGACGACAA 275
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ATTTAGGATTTCAGGGAAAAGTGTCCAAGCTTTCAGTGTTGGAGCAGGTATGGACGACAA 240

Qy     276 AGGCGACCCGAGCAATGAGGAGGCACCTAAGGCCATCAAACCCACCAGCAAAGAGTTCAG 335
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AGGCGACCCGAGCAATGAGGAGGCACCTAAGGCCATCAAACCCACCAGCAAAGAGTTCAG 300

Qy     336 GAAAACATGGGGTTTTCGAAGGACCACTATCGCCAAGCGAGAGGGCGCAGGGGACGCGGA 395
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GAAAACATGGGGTTTTCGAAGGACCACTATCGCCAAGCGAGAGGGCGCAGGGGACGCGGA 360

Qy     396 GGCTGACCCACTGGAGCCGCCACCCCCACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAG 455
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GGCTGACCCACTGGAGCCGCCACCCCCACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAG 420

Qy     456 TGGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCCG 515
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TGGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCCG 480

Qy     516 CGGCAGGAGGAGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCGC 575
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CGGCAGGAGGAGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCGC 540

Qy     576 CACAGACGCCGAGACAGCCTCCGAGGGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGG 635
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CACAGACGCCGAGACAGCCTCCGAGGGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGG 600

Qy     636 CCCCCAGTCTGCTTCCACAGCTGTGAAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAGG 695
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 CCCCCAGTCTGCTTCCACAGCTGTGAAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAGG 660

Qy     696 AGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGGCCTGACCTTGAAAGAGCT 755
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 AGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGGCCTGACCTTGAAAGAGCT 720

Qy     756 TCAGAATCGCCTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGAT 815
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 TCAGAATCGCCTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGAT 780

Qy     816 CCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGA 875
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 CCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGA 840

Qy     876 GGCCAGTGACACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGGG 935
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GGCCAGTGACACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGGG 900

Qy     936 GGTTGTGTCCCAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCA 995
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 GGTTGTGTCCCAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCA 960

Qy     996 GGACATCAAAGATGAGGAGCCTGGAGACTTGGGCCGACCGAAGCCTGAATGTGAGGGTTA 1055
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GGACATCAAAGATGAGGAGCCTGGAGACTTGGGCCGACCGAAGCCTGAATGTGAGGGTTA 1020

Qy    1056 CGACCCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTTG 1115
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CGACCCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTTG 1080

Qy    1116 CTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAGG 1175
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAGG 1140

Qy    1176 GAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACTGCACCATTCTGCAAGT 1235
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 GAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACTGCACCATTCTGCAAGT 1200

Qy    1236 GCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGA 1295
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 GCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGA 1260

Qy    1296 TGCTGATGGCACCGATTGTACAAGTATAGGAACAATAGAGCAGAAGTCTAGCGAAGACCA 1355
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 TGCTGATGGCACCGATTGTACAAGTATAGGAACAATAGAGCAGAAGTCTAGCGAAGACCA 1320

Qy    1356 AGGGATAAAGGGTAGAATTGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGAT 1415
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 AGGGATAAAGGGTAGAATTGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGAT 1380

Qy    1416 CTTCCAGCCTGTGATAGAGGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCA 1475
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 CTTCCAGCCTGTGATAGAGGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCA 1440
```

```
Qy    1476 CGTGGCGCAGCCCGACTCGGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAGC 1535
Db    1441 CGTGGCGCAGCCCGACTCGGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAGC 1500

Qy    1536 GACAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAA 1595
Db    1501 GACAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAA 1560

Qy    1596 GATGAAGCCAGAGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTC 1655
Db    1561 GATGAAGCCAGAGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTC 1620

Qy    1656 TTCTGTGCACAAGAGACCAGCTCCAG-AAAAAAAAGAGACCACAGTGAAGAAGGCAGTGG 1714
Db    1621 TTCTGTGCACAAGAGACCAGCTCCAGAAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGG 1680

Qy    1715 TGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGT 1774
Db    1681 TGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGT 1740

Qy    1775 CGTGGGCGAGCGATCACAATTACAATGCAGTAAAGCCAGAAAAGACTGCTGCTCCCTCGC 1834
Db    1741 CGTGGGCGAGCGATCACAATTACAATGCAGTAAAGCCAGAAAAGACTGCTGCTCCCTCGC 1800

Qy    1835 CGTCACTGTTGTATAAATGTATGTATCACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTT 1894
Db    1801 CGTCACTGTTGTATAAATGTATGTATCACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTT 1860

Qy    1895 CTTTCTGGATAGCCATCCCCTGGGCCTGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAG 1954
Db    1861 CTTTCTGGATAGCCATCCCCTGGGCCTGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAG 1920

Qy    1955 CTGATCACAGACACCGGCTGCACCATCAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTC 2014
Db    1921 CTGATCACAGACACCGGCTGCACCATCAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTC 1980

Qy    2015 CTGCTGCCCTGTGTCCATCCCTAGTCTGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTG 2074
Db    1981 CTGCTGCCCTGTGTCCATCCCTAGTCTGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTG 2040

Qy    2075 TAAACCTCACTGGTGAACGTTCACCTTAATGATTGATTCTTTAATCTCTGTTTTCACTCT 2134
Db    2041 TAAACCTCACTGGTGAACGTTCACCTTAATGATTGATTCTTTAATCTCTGTTTTCACTCT 2100

Qy    2135 CAGGCTCTGGTAAGTATTTGTATTCTCTTCATCCCAGTCTGATTGCATAGCCACACTGCC 2194
Db    2101 CAGGCTCTGGTAAGTATTTGTATTCTCTTCATCCCAGTCTGATTGCATAGCCACACTGCC 2160

Qy    2195 CGGCACGCCACATCCACCCCTGTCTGCACATGAGTTGTTCTGACAACAGCGCTGTATACG 2254
Db    2161 CGGCACGCCACATCCACCCCTGTCTGCACATGAGTTGTTCTGACAACAGCGCTGTATACG 2220

Qy    2255 CTTCAGTTTTTCCACATTGTCCACGGCCAGCACATGAAAGCATCACTTCTTTTTTATGTT 2314
Db    2221 CTTCAGTTTTTCCACATTGTCCACGGCCAGCACATGAAAGCATCACTTCTTTTTTATGTT 2280

Qy    2315 GTGGGAATCTTTGCAAGTTAGTGTTGCATCTGATTTTCAGGTGTACATTTATTTTTGACT 2374
Db    2281 GTGGGAATCTTTGCAAGTTAGTGTTGCATCTGATTTTCAGGTGTACATTTATTTTTGACT 2340

Qy    2375 GGGCAGATAGGGGATTTTTTTTTTTTTCCATGTCCGATTCACACGCTACACACCCACATG 2434
Db    2341 GGGCAGATAGGGGA--TTTTTTTTTTTCCATGTCCGATTCACACGCTACACACCCACATG 2398

Qy    2435 AACACATTCGAACTTCGAAGG-CACACACTCCTGCTTCATAGGCCCCACGGTAAGTGAGT 2493
Db    2399 AACACATTCGAACTTCGAAGGCCACACACTCCTGCTTCATAGGCCCCACGGTAAGTGAGT 2458

Qy    2494 TCACACCTAGAACACTGTCCTGACCGCAGGACGCGTGCCTTGGACTTGGTATTCTACATG 2553
Db    2459 TCACACCTAGAACACTGTCCTGACCGCAGGACGCGTGCCTTGGACTTGGTATTCTACATG 2518

Qy    2554 TGACTGGCTTTCTTGCCCTCGTCTCTTGAATGTTTAGACTCTTAAGATCATATCCTG 2610
Db    2519 TGACTGGCTTTCTTGCCCTCGTCTCTTGAATGTTTAGACTCTTAAGATCATATCCTG 2575
```

RESULT 2
AK042474

```
LOCUS       AK042474                4619 bp    mRNA    linear   HTC 19-SEP-2003
DEFINITION  Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched
            library, clone:A630095A07 product:death inducer-obliterator-1, full
            insert sequence.
ACCESSION   AK042474
VERSION     AK042474.1  GI:26335138
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 4619)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
```

```
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
            Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
            Trust/MRC building Addenbrookes Hospital Cambridge) whose
            assistance we gratefully acknowledge.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES                Location/Qualifiers
     source             1. .4619
                        /organism="Mus musculus"
                        /mol_type="mRNA"
                        /strain="C57BL/6J"
                        /db_xref="FANTOM_DB:A630095A07"
                        /db_xref="MGI:2407519"
                        /db_xref="taxon:10090"
                        /clone="A630095A07"
                        /tissue_type="thymus"
                        /clone_lib="RIKEN full-length enriched mouse cDNA library"
                        /dev_stage="3 days neonate"
     CDS                299. .2143
                        /note="unnamed protein product; death
                        inducer-obliterator-1 (MGD|MGI:1344352, GB|AJ238332,
                        evidence: BLASTN, 99%, match=2599)
                        putative"
                        /codon_start=1
                        /protein_id="BAC31270.1"
                        /db_xref="GI:26335139"
                        /translation="MDDKGHLSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTE
                        VDPSEQQPQQHNLSLRRSGRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTV
                        TDVETASEGSVESSSEIRSGPVSDSLGKEHPASSEKAKGGEEEEDTSDSDSDGLTLKE
                        LQNRLRRKREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSAEQDRPLCKQEPEA
                        SQGPVSQSETDDIENQLEGKATQGNTEENPREAGKPKPECEVYDPNALYCICRQPHNN
                        RFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATDEQD
                        SGCRSVGADGTDCTSIGTIEQKSGEDQGIKGRIEKAANPSGKKKLKIFQPVVEAPGAP
                        KCIGPGCSSVAQPDSVYCSNDCILKHAAATMRFLSSGKEQKTKPKEKVKTKPEKFSLP
                        KCSVQVGIKISSVHKRLASEKRENPVKKVMLASRSETSGKEAACESSTPSWASDHNYN
                        AVKPEKPEKPTALSPTLLSKCTYHPKAGFPGPSHHLGGCLGLSRTRVLGVLVLIVASS
                        SLPARSRYQDASGPQVFLPSLWSLSGWFLKSCVGLMLEAISYFSFRPW"
ORIGIN

  Query Match             39.3%;  Score 1025.2;  DB 11;  Length 4619;
  Best Local Similarity   71.9%;  Pred. No. 1.4e-243;
  Matches 1495;  Conservative    0;  Mismatches  533;  Indels   50;  Gaps   10;

Qy         21 TCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGCCTGCGGAGCTTACTCCACGG 80
Db         68 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 127

Qy         81 GAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTGTTACTCGTGAACAGTG 140
Db        128 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 187

Qy        141 GCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCCAGAGGTTTCGTCTGCC 200
Db        188 ACTGACAACAGTGGGGTGAG--GCTTGGCCGTCTGCTTGACC------------TGGCCC 233

Qy        201 AGGGTTTTTGGTTGTATTTAGGATTT-CAGGGAAAAGTGTCCAAGCTTTCAGTGTTGGAG 259
Db        234 CAGGTCTATAATTTTATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGA 293

Qy        260 CAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTAAGGCCATCAAACCCA 319
Db        294 CAGGTATGGATGATAAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCA 353

Qy        320 CCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTATCGCCAAGCGAGAGG 379

Db        354 CCAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGG 413

Qy        380 GCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCACAGCAGCAGCTGGGCC 439
Db        414 GTGCAGGAGACACGGAGGTGGACCCCAGTGAGCAGCAACC------ACAGCAGCATAACC 467

Qy        440 TGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGA 499
Db        468 TCTCCCTGCGCCGCAGTGGACGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTA 527

Qy        500 CCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGC 559
Db        528 CCACGGTTCGGCGCCGAGGGAAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGC 587

Qy        560 CCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCAGCGTGGAAAGCGCTT 619
Db        588 CCACATCTTCCACAGTCACTGATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTT 647

Qy        620 CTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGGAACGACCAGCCTCTT 679
Db        648 CTGAGATCAGAAGTGGCCCTGTATCTGACTCCTTA---GGGAAAGAACATCCTGCCTCTT 704

Qy        680 CTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGGCC 739
Db        705 CTGAAAAGGCAAAAGGAGGTGAAGAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCC 764

Qy        740 TGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGA 799
Db        765 TTACGTTGAAGGAACTTCAGAACCGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGA 824

Qy        800 GGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCG 859
Db        825 GGTCCCTGAGAGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCG 884

Qy        860 AGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGC 919
Db        885 AAACTGGGAGTGTCCAAATAGGCAGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGC 944

Qy        920 CCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGG 979
Db        945 CTGAGGCTAGTCAGGGACCAGTGTCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGG 1004

Qy        980 AGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACTTGGGCCGACCGAAGC 1039
Db       1005 AAGGGAAGGCGACTCAGGGAAATACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGC 1064

Qy       1040 CTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACA 1099
Db       1065 CTGAGTGTGAGGTTTACGACCCCAATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACA 1124

Qy       1100 ACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCA 1159
Db       1125 ACAGGTTTATGATCTGCTGTGATCGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTA 1184

Qy       1160 TTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACT 1219
Db       1185 TTTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATT 1244

Qy       1220 GCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTA 1279
Db       1245 GCACCATTTTGCAAGTGCAGGATGAGACAAACGGTAGCGCCACCGATGAGCAGGACTCTG 1304

Qy       1280 AATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAGGAACAATAGAGCAGA 1339
Db       1305 GGTGCAGATCTGTGGGTGCTGATGGCACAGACTGCACAAGCATAGGGACAATAGAGCAGA 1364

Qy       1340 AGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTGCAAATCCAAGTGGCA 1399
Db       1365 AGTCCGGAGAAGACCAGGGCATAAAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCA 1424

Qy       1400 AGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGGTGCCTCAAAATGTATTG 1459
```

```
Db       1425 AGAAAAAACTCAAGATATTCCAGCCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTG 1484

Qy       1460 GCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTGCAGTAATGACTGTATCC 1519

Db       1485 GCCCTGGGTGTTCCAGTGTAGCACAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTC 1544

Qy       1520 TCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGC 1579

Db       1545 TCAAACACGCAGCAGCTACCATGAGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAAC 1604

Qy       1580 CTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGG 1639

Db       1605 CCAAGGAAAAGGTCAAGACGAAGCCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGG 1664

Qy       1640 CAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAAAAAAAGAGACCACAG 1699

Db       1665 TGGGGATTAAAATCTCTTCTGTGCACAAGAGACTAGCGTCAGAGAAAAGGGAAAACCCAG 1724

Qy       1700 TGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTG 1759

Db       1725 TGAAGAA---AGTGATGCTGGCTTCCAGGAGTGAGACTTCTGGGAAGGAGGCAGCCTGTG 1781

Qy       1760 AGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAGTAAAGCCAGAAAAGA 1819

Db       1782 AGAGCAGCACACCATCCTGGGCAAGTGACCACAACTACAATGCTGTGAAGCCAGAGAAGC 1841

Qy       1820 CTG---------CTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACCTAGGGG 1870

Db       1842 CAGAGAAGCCCACTGCACTCTCGCCCACCCTATTGAGTAAATGTACGTATCACCCAAAGG 1901

Qy       1871 TTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCCTGTCCAGGAC 1930

Db       1902 CTGGCTTCCCAGGCCCCTCCCATCATCT-----GGGTGGCTGCCTGGGGCTGTCTAGGAC 1956

Qy       1931 TGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCAGCGGGAAG 1990

Db       1957 CAG-----AGTCCTGGGTGTTCTGGTGCTGATAGTAGCCAGCAGC----TCACTGCCAGC 2007

Qy       1991 CAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCCTGTGTCCATCCCTAGTCTGTCAGGACT 2050

Db       2008 CAGAAGCAGATACCAAGATGCCTCTGGACCCCAGGTGTTCCTGCCTAGCCTGTGGAGCCT 2067

Qy       2051 TCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGT 2088

Db       2068 CTCTGGGTGGTTCCTAAAGAGCTGTGTAGGCCTCATGT 2105


RESULT 3
AL578222/c
LOCUS       AL578222               1086 bp    mRNA    linear   EST 01-JUN-2003
DEFINITION  AL578222 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
            cDNA clone CS0DK002YJ13 3-PRIME, mRNA sequence.
ACCESSION   AL578222
VERSION     AL578222.2  GI:31316440
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1086)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On Feb 16, 2001 this sequence version replaced gi:12942093.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 9799.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DK002CE07NP1&cluster=9799.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DK002CE07NP1.
FEATURES             Location/Qualifiers
     source          1. .1086
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DK002YJ13"
                     /cell_type="HELA CELLS COT 25-NORMALIZED"
                     /cell_line="HELA"
                     /clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             38.6%;  Score 1007.2;  DB 9;  Length 1086;
  Best Local Similarity   97.5%;  Pred. No. 2.8e-239;
  Matches 1047;  Conservative    5;  Mismatches   18;  Indels    4;  Gaps    3;

Qy       1047 TGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACAACAGGTT 1106

Db       1072 TGTGAGGTTACGACCCAACGCCCTGTAT--GCWTTGCCSCCAGCCTCACAACAACA-GTT 1016

Qy       1107 TATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCATTTCTGA 1166

Db       1015 TATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCATTTCTGA 956

Qy       1167 GGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACTGCACCAT 1226

Db        955 GGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACTGCACCAT 896

Qy       1227 TCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTAAATGGAG 1286

Db        895 TCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTAAATGGAG 836

Qy       1287 ACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAGGAACAATAGAGCAGAAGTCTAG 1346

Db        835 ACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAGGAACAATAGAGCAGAAGTCTAG 776

Qy       1347 CGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAA 1406

Db        775 CGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAR 716

Qy       1407 ACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGGTGCCTCAAAATGTATTGGCCCCGG 1466

Db        715 ACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGGTGCCTCAAAATGTATTGGCCCCGG 656

Qy       1467 GTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTGCAGTAATGACTGTATCCTCAAACA 1526

Db        655 GTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTGCAGTAATGACTGTATCCTCAAACA 596

Qy       1527 CGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGA 1586

Db        595 CGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGA 536

Qy       1587 AAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGGCAGGTAT 1646

Db        535 AAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGGCAGGTAT 476

Qy       1647 TAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAAAAAAAGAGACCACAGTGAAGAA 1706

Db        475 TAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAAAAAAAGAGACCACAGTGAAGAA 416

Qy       1707 GGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAG 1766

Db        415 GGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAG 356

Qy       1767 CACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAGTAAAGCCAGAAAAGACTGCTGC 1826

Db        355 CACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAGTAAAGCCAGAAAAGACTGCTGC 296
```

```
Qy       1827 TCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACCTAGGGGTTGGCCTCCTGGACCC 1886
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        295 TCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACCTAGGGGTTGGCCTCCTGGACCC 236

Qy       1887 CTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCCTGTCCAGGACTGGGAGTTGCAGCTTT 1946
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 CTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCCTGTCCAGGACTGGGAGTTGCAGCTTT 176

Qy       1947 GTGTTAAGCTGATCACAGACACCGGCTGCACCATCAGCGGGAAGCAGAGCCCATGTCCAG 2006
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        175 GTGTTAAGCTGATCACAGACACCGGCTGCACCATCAGCGGGAAGCAGAGCCCATGTCCAG 116

Qy       2007 GATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTCTGTCAGGACTTCCTGTCACTGTTTTC 2066
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        115 GATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTCTGTCAGGACTTCCTGTCACTGTTTTC 56

Qy       2067 CAAAGCTGTAAACCTC-ACTGGTGAACGTTCACCTTAATGATTGATTCTTTAAT 2119
             |||||||||||||||| |||||||||||||||||||||    :    || |:|
Db         55 CAAAGCTGTAAACCTCNACTGGTGAACGTTCACCTTAATGNNNDGANTTTAAMT 2
```

RESULT 4
BC060249

LOCUS	BC060249 4651 bp mRNA linear HTC 19-NOV-2003
DEFINITION	Mus musculus death inducer-obliterator 1, mRNA (cDNA clone IMAGE:6826043), containing frame-shift errors.
ACCESSION	BC060249
VERSION	BC060249.1 GI:37805411
KEYWORDS	HTC.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 4651)
AUTHORS	Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE	22388257
PUBMED	12477932
REFERENCE	2 (bases 1 to 4651)
AUTHORS	Strausberg,R.
TITLE	Direct Submission
JOURNAL	Submitted (20-OCT-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk Email: cgapbs-r@mail.nih.gov Tissue Procurement: Dr. Jim Lin, University of Iowa cDNA Library Preparation: M. Bento Soares, University of Iowa cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada info@bcgsc.bc.ca Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAK Plate: 126 Row: k Column: 20 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein This clone has the following problem: frame shifted.

FEATURES	Location/Qualifiers
source	1. .4651
	/organism="Mus musculus"
	/mol_type="mRNA"
	/strain="C57BL/6"
	/db_xref="taxon:10090"
	/clone="IMAGE:6826043"
	/tissue_type="Brain, mouse, 13.5,14.5,16.5,17.5 dpc"
	/clone_lib="NIH_BMAP_FYO"
	/lab_host="DH10B"
	/note="Vector: pYX-ASC"

ORIGIN

Query Match 38.4%; Score 1002.6; DB 11; Length 4651;
Best Local Similarity 73.5%; Pred. No. 5.8e-238;
Matches 1414; Conservative 0; Mismatches 474; Indels 37; Gaps 9;

```
Qy        175 GCTTGGACCCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTT-CAGGGAA 233
              |||||| |     |  | |   || ||| | |  || ||| |||||||| |||  ||
Db        187 GCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTTTATGTAGGATTTCCAGCCAA 246

Qy        234 AAGTGTCCAAGCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGA 293
              | || ||||||||||||||| ||||  |||||||||| || |||||  ||| |||||||||
Db        247 AGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGATAAAGGGCACCTGAGCAATGA 306

Qy        294 GGAGGCACCTAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCG 353
              ||| ||||| |||||||||||||||| || ||||||||||||| ||||||||||| ||||
Db        307 GGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCG 366

Qy        354 AAGGACCACTATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCC 413
              ||| ||||| || ||||| || ||||| ||||| ||| ||||||  |||||    ||||
Db        367 AAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACGGAGGTGGACCCCAGTGAGCA 426

Qy        414 GCCACCCCCACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCG 473
              || |||       |||||||    ||| ||||||||  ||||||||  |||| || || ||
Db        427 GCAACC------ACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGACGGCAACCAAAACG 480

Qy        474 CACTGAGCGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCC 533
              |||||| | || ||  |||| || ||||  |  ||||||||  || |  | ||   ||||
Db        481 TACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGGAAAAAGAATGTGCC 540

Qy        534 TGTCTCCCTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGC 593
              || |||||||||||||  ||||||||| || | | | ||| || |  ||||||||
Db        541 GGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACTGATGTGGAGACAGC 600

Qy        594 CTCCGAGGGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCAC 653
              ||||||||  |||||  ||||||  ||||||||  ||||||  |||||     ||||  |||
Db        601 TTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCTGTATCTGACTCCTT 660

Qy        654 AGCTGTGAAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGA 713
              |   | |||  |||||  ||||||||||||||||  ||||||||  || ||   | || ||
Db        661 A---GGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAAGAGGAAGAAGA 717

Qy        714 CACCTCCGATAGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAG 773
              |||||| || |||||||| ||||||||  || ||||| || ||||||||  ||||||||  ||
Db        718 CACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAACCGCCTTCGGAG 777
```

```
Qy        774 GAAGCGGGAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAA 833
Db        778 AAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAATCGCCTGAGGAA 837

Qy        834 GAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGA 893
Db        838 GAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGCAGTGCCGAGCA 897

Qy        894 GGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGG 953
Db        898 GGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTGTCCCAGTCAGA 957

Qy        954 GAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGA 1013
Db        958 GACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAATACAGAGGAAAA 1017

Qy       1014 GCCTGGAGACTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTA 1073
Db       1018 CCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCCAATGCCCTGTA 1077

Qy       1074 TTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGA 1133
Db       1078 CTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGATCGGTGTGAGGA 1137

Qy       1134 ATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAA 1193
Db       1138 GTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAA 1197

Qy       1194 TGGGGAAGACTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTC 1253
Db       1198 CGGGGAAGACTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGATGAGACAAACGG 1257

Qy       1254 AGAAACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTG 1313
Db       1258 TAGCGCCACCGATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGATGGCACAGACTG 1317

Qy       1314 TACAAGTATAGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAAT 1373
Db       1318 CACAAGCATAGGGACAGTAGAGCAGAAGTCCGGAGAAGACCAGGGCATAAAGGGTAGGAT 1377

Qy       1374 TGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGA 1433
Db       1378 TGAGAAGGCAGCAAACCCCAGCGGCAAGAAAAAACTCAAGATATTCCAGCCTGTCGTAGA 1437

Qy       1434 GGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTC 1493
Db       1438 GGCTCCTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCACAGCCTGACTC 1497

Qy       1494 GGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAG 1553
Db       1498 TGTGTATTGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATGAGATTTCTAAG 1557

Qy       1554 CTCAGGTAAAGAAC-AGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGC 1612
Db       1558 TTCAGGTAAAGAACAAAAAAACAAAACCCAAGGAAAAGGTCAAGACGAAGCCAGAAAAGT 1617

Qy       1613 CCAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGAC 1672
Db       1618 TCAGTCTTCCAAAATGCAGTGTTCAGGTGGGGATTAAAATCTCTTCTGTGCACAAGAGAC 1677

Qy       1673 CAGCTCCAGAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTG 1732
Db       1678 TAGCGTCAGAGAAAAGGGAAAACCCAGTGAAGAA---AGTGATGCTGGCTTCCAGGAGTG 1734

Qy       1733 AAGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACA 1792
Db       1735 AGACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGCACACCATCCTGGGCAAGTGACCACA 1794

Qy       1793 ATTACAATGCAGTAAAGCCAGAAAAGACTG---------CTGCTCCCTCGCCGTCACTGT 1843
Db       1795 ACTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAGCCCACTGCACTCTCGCCCACCCTAT 1854

Qy       1844 TGTATAAATGTATGTATCACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGA 1903
Db       1855 TGAGTAAATGTACGTATCACCCAAAGGCTGGCTTCCCAGGCCCCTCCCATCATCT----- 1909

Qy       1904 TAGCCATCCCCTGGGCCTGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACA 1963
Db       1910 GGGTGGCTGCCTGGGGCTGTCTAGGACCAG-----AGTCCTGGGTGTTCTGGTGCTGATA 1964

Qy       1964 GACACCGGCTGCACCATCAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCC 2023
Db       1965 GTAGCCAGCAGC----TCACTGCCAGCCAGAAGCAGATACCAAGATGCCTCTGGACCCCA 2020

Qy       2024 TGTGTCCATCCCTAGTCTGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCA 2083
Db       2021 GGTGTTCCTGCCTAGCCTGTGGAGCCTCTCTGGGTGGTTCCTAAAGAGCTGTGTAGGCCT 2080

Qy       2084 CTGGT 2088
Db       2081 CATGT 2085
```

RESULT 5
AL555810
LOCUS AL555810 1093 bp mRNA linear EST 31-MAY-2003
DEFINITION AL555810 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens cDNA clone CS0DK002YJ13 5-PRIME, mRNA sequence.
ACCESSION AL555810
VERSION AL555810.2 GI:31277616
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1093)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 15, 2001 this sequence version replaced gi:12897895.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 9799.r For more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DK002CE07QP1&cluster=9799.r. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Paraday Avenue Genoscope sequence ID : CS0DK002CE07QP1.
FEATURES Location/Qualifiers
source 1. .1093
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK002YJ13"
/cell_type="HELA CELLS COT 25-NORMALIZED"
/cell_line="HELA"
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

Query Match 36.7%; Score 957.2; DB 9; Length 1093;
Best Local Similarity 97.8%; Pred. No. 7.8e-227;
Matches 1011; Conservative 3; Mismatches 11; Indels 9; Gaps 4;

```
Qy         41 GGCTGCGAGACCCTAGAGGCCTGCG-----GAGCTTACTCCACGGGAACAGCCTCTAGAT 95
Db         57 GGATGCGAGACCCTAGAGGCCTGCGGCTGCGGACTTACTCCACGGGAACAGCCTCTAGAT 116
```

```
Qy         96 AATCTGAGTTGTTGAAAATACGAAGCCTGTTACTCGTGAACAGTGGCTGACAACAGTGTT 155
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        117 AATCTGAGTTGTTGAAAATACGAAGCCTGTTACTCGTGAACAGTGGCTGACAACAGTGTT 176

Qy        156 GTTGTGAGCCTGGCTGTCTGCTTGGACCCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGT 215
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        177 GTTGTGAGCCTGGCTGTCTGCTTGGACCCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGT 236

Qy        216 ATTTAGGATTTCAGGGAAAAGTGTCCAAGCTTTCAGTGTTGGAGCAGGTATGGACGACAA 275
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        237 ATTTAGGATTTCAGGGAAAAGTGTCCAAGCTTTCAGTGTTGGAGCAGGTATGGACGACAA 296

Qy        276 AGGCGACCCGAGCAATGAGGAGGCACCTAAGGCCATCAAACCCACCAGCAAAGAGTTCAG 335
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        297 AGGCGACCCGAGCAATGAGGAGGCACCTAAGGCCATCAAACCCACCAGCAAAGAGTTCAG 356

Qy        336 GAAAACATGGGGTTTTCGAAGGACCACTATCGCCAAGCGAGAGGGCGCAGGGGACGCGGA 395
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        357 GAAAACATGGGGTTTTCGAAGGACCACTATCGCCAAGCGAGAGGGCGCAGGGGACGCGGA 416

Qy        396 GGCTGACCCACTGGAGCCGCCACCCCCACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAG 455
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        417 GGCTGACCCACTGGAGCCGCCACCCCCACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAG 476

Qy        456 TGGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCCG 515
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        477 TGGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCCG 536

Qy        516 CGGCAGGAGGAGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCGC 575
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        537 CGGCAGGAGGAGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCGC 596

Qy        576 CACAGACGCCGAGACAGCCTCCGAGGGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGG 635
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        597 CACAGACGCCGAGACAGCCTCCGAGGGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGG 656

Qy        636 CCCCCAGTCTGCTTCCACAGCTGTGAAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAGG 695
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        657 CCCCCAGTCTGCTTCCACAGCTGTGAAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAGG 716

Qy        696 AGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGGCCTGACCTTGAAAGAGCT 755
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        717 AGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGGCCTGACCTTGAAAGAGCT 776

Qy        756 TCAGAATCGCCTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGAT 815
              |||||||||||||||||||||||||||||||||||||||||||||||||  || ||||
Db        777 TCAGAATCGCCTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGAGGCCC--TGAAAGGAT 834

Qy        816 CCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGA 875
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        835 CCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGA 894

Qy        876 GGCCAGTGACACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAACGATCA-GG 934
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        895 GGCCAGTGACACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGGG 954

Qy        935 GGGTTGTGTCCCAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTC 994
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        955 GGGTTGTGTCCCAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTC 1014

Qy        995 AGGACATCAAAGATGAGGAGCCTGGAGACTTGGGCCGACCGAAGCCTGAATGTGAGGGTT 1054
              |||||||||||||||||||||||||||||| |||||||| |:|||||||||:|| |||
Db       1015 AGGACATCAAAGATGAGGAGCCTGGAGACTTTGGCCGACCG-ARCCTGAATGTKAGAGTT 1073

Qy       1055 ACGACCCCAACGCC 1068
              |||||||:|   ||
Db       1074 ACGACCCMACGCCC 1087
```

RESULT 6
AL580669

```
LOCUS       AL580669              1033 bp    mRNA    linear   EST 01-JUN-2003
DEFINITION  AL580669 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
            Homo sapiens cDNA clone CS0DJ007YD06 5-PRIME, mRNA sequence.
ACCESSION   AL580669
VERSION     AL580669.2  GI:31318936
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1033)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On Feb 16, 2001 this sequence version replaced gi:12946913.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 9799.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DJ007DB03QP1&cluster=9799.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DJ007DB03QP1.
FEATURES             Location/Qualifiers
     source          1. .1033
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DJ007YD06"
                     /cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
                     /cell_line="JURKAT"
                     /clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
                     10-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             34.9%;  Score 910.6;  DB 9;  Length 1033;
  Best Local Similarity   97.0%;  Pred. No. 3.1e-215;
  Matches  934;  Conservative   14;  Mismatches    9;  Indels    6;  Gaps    2;

Qy         33 GGGGAAATGGCTGCGAGACCCTAGAGGCCTGCGGAG-----CTTACTCCACGGGAACAGC 87
              ||||||||||||||||||||||:||||||||     |||||||||||||||||
Db         67 GGGGAAATGGCTGCGAGACCCTAGAVGCCTGCGGCTGCGGACTTACTCCACGGGAACAGC 126

Qy         88 CTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTGTTACTCGTGAACAGTGGCTGACA 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 CTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTGTTACTCGTGAACAGTGGCTGACA 186

Qy        148 ACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCCAGAGGTTTCGTCTGCCAGGGTTT 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        187 ACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCCAGAGGTTTCGTCTGCCAGGGTTT 246

Qy        208 TTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCAAGCTTTCAGTGTTGGAGCAGGTATG 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        247 TTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCAAGCTTTCAGTGTTGGAGCAGGTATG 306

Qy        268 GACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTAAGGCCATCAAACCCACCAGCAAA 327
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        307 GACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTAAGGCCATCAAACCCACCAGCAAA 366

Qy        328 GAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTATCGCCAAGCGAGAGGGCGCAGGG 387
              :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        367 RAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTATCGCCAAGCGAGAGGGCGCAGGG 426

Qy        388 GACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCACAGCAGCAGCTGGGCCTGTCCCTG 447
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        427 GACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCACAGCAGCAGCTGGGCCTGTCCCTG 486
```

```
Qy        448 CGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGACCATTGCG 507
Db        487 CGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGACCATTGCG 546

Qy        508 CGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGCCCACGTCC 567
Db        547 CGGCGCCGCGGCAGGAGGAGCATTCCTTTCTCCCTGGAGGATTCTGGTGAGCCCACGTCC 606

Qy        568 TGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCAGCGTGGAAAGCGCTTCTGAGACC 627
Db        607 TGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCAGCGTGGAAAGCGCTTCTGAGACC 666

Qy        628 AGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGGAACGACCAGCCTCTTCTGAAAAG 687
Db        667 AGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGGAACGACCAGCCTCTTCTGAAAAG 726

Qy        688 GTGAAAGGAGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGGCCTGACCTTG 747
Db        727 GTGAAAGGAGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGGCCTGACCTTG 786

Qy        748 AAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGAGGCCCCTG 807
Db        787 AAARAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGAGGCCCCTG 846

Qy        808 AAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTG 867
Db        847 AAAGGAATCCAGAGTCGCCTGCGGAAARAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTG 906

Qy        868 GGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAAC 927
Db        907 GGCTCCGAGGCCARTGACACTATGGAGGGCGT-CYKSCCARTAARCAGGAGCCCGARAAC 965

Qy        928 GATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAG 987
Db        966 GATCAGGGGGTTGTKTCCCAGGCTGGGAAAAAWGAMAGAGAAAATAAGTTGGAGGGAAAG 1025

Qy        988 GCG 990
Db       1026 GCG 1028
```

```
RESULT 7
AL576399/c
LOCUS       AL576399               1201 bp    mRNA    linear   EST 01-JUN-2003
DEFINITION  AL576399 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI075YL03 3-PRIME, mRNA sequence.
ACCESSION   AL576399
VERSION     AL576399.2  GI:31314685
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On Feb 16, 2001 this sequence version replaced gi:12938505.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 9799.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI075CF02NP1&cluster=9799.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI075CF02NP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI075YL03"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
```

Query Match 34.7%; Score 905.6; DB 9; Length 1201;
Best Local Similarity 93.8%; Pred. No. 5.6e-214;
Matches 972; Conservative 19; Mismatches 39; Indels 6; Gaps 5;

```
Qy       1568 AGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGAAAT 1627
Db       1073 AWAAGAMRAAGCMAAGCCTAAGVAAAGATGARGATAAGCCAGAGAGCCCAGTTTCCGAAA 1014

Qy       1628 GCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAAAAA 1687
Db       1013 TGCGGGCTCAGGCAGKTW-TAAAATYTYTTTYTGTGCACAAGAGACCAGCTCCA-RAAAAA 956

Qy       1688 AAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGG 1747
Db        955 AAGAGACCACAKKRAGRAAGGMAKTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGG 896

Qy       1748 AAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAGTAA 1807
Db        895 AAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGRGCGATCACAATTACAATGCAGTAA 836

Qy       1808 AGCCAG-AAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACCTA 1866
Db        835 AGCCAGAAAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACCTA 776

Qy       1867 GGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCCTGTCCA 1926
Db        775 GGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCCTGTCCA 716

Qy       1927 GGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCAGCGG 1986
Db        715 GGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCAGCGG 656

Qy       1987 GAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTCTGTCAG 2046
Db        655 GAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTCTGTCAG 596

Qy       2047 GACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGTGAACGTTCACCTTAATGA 2106
Db        595 GACTKCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGTGAACGTTCACCTTAATGA 536

Qy       2107 TTGATTCTTTAATCTCTGTTTTCACTCTCAGGCTCTGGTAAGTATTTGTATTCTCTTCAT 2166
Db        535 TTGATTCTTTAATCTCTGTTTTCACTCTCAGGCTCTGGTAAGTATTCGTATTCTCTTCAT 476

Qy       2167 CCCAGTCTGATTGCATAGCCACACTGCCCGGCACGCCACATCCACCCCTGTCTGCACATG 2226
Db        475 CCCAGTCTGATTGCATAGCCACACTGCCCGGCACGCCACATCCACCCCTGTCTGCACATG 416

Qy       2227 AGTTGTTCTGACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGGCCAGCA 2286
Db        415 AGTTGTTCTGACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGGCCAGCA 356

Qy       2287 CATGAAAGCATCACTTCTTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTGCATCTG 2346
Db        355 CATGAAAGCATCACTTCTTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTGCATCTG 296

Qy       2347 ATTTTCAGGTGTACATTTATTTTTGACTGGGCAGATAGGGGATTTTTTTTTTTTTCCATG 2406
Db        295 ATTTTCAGGTGTACATTTATTTTTGACTGGGCAGATAGGGGA--TTTTTTTTTTTCCATG 238

Qy       2407 TCCGATTCACACGCTACACACACCCACATGAACACATTCGAACTTCGAAGG-CACACACTCC 2465
```

```
Db        237 TCCGATTCACACGCTACACACCCACATGAACACATTCGAACTTCGAAGGCCACACACTCC 178

Qy       2466 TGCTTCATAGGCCCCACGGTAAGTGAGTTCACACCTAGAACACTGTCCTGACCGCAGGAC 2525
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        177 TGCTTCATAGGCCCCACGGTAAGTGAGTTCACACCTAGAACACTGTCCTGACCGCAGGAC 118

Qy       2526 GCGTGCCTTGGACTTGGTATTCTACATGTGACTGGCTTTCTTGCCCTCGTCTCTTGAATG 2585
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        117 GCGTGCCTTGGACTTGGTATTCTACATGTGACTGGCTTTCTTGCCCTCGTCTCTTGAATG 58

Qy       2586 TTTAGACTCTTAAGAT 2601
              ||||||||||||||||
Db         57 TTTAGACTCTTAAGAT 42


RESULT 8
BX380892
LOCUS       BX380892               1010 bp    mRNA    linear   EST 08-MAY-2003
DEFINITION  BX380892 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI058YJ17 5-PRIME, mRNA sequence.
ACCESSION   BX380892
VERSION     BX380892.1  GI:30439629
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1010)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 9799.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI058CE09QP1&cluster=9799.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI058CE09QP1.
FEATURES             Location/Qualifiers
     source          1. .1010
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI058YJ17"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             34.5%;  Score 900.2;  DB 13;  Length 1010;
  Best Local Similarity   98.6%;  Pred. No. 1.2e-212;
  Matches  934;  Conservative    5;  Mismatches    4;  Indels    4;  Gaps    3;

Qy       1599 GAAGCCAGAGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTC 1658
              || :|||||:||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 GATKCCAGAKAAGCCCAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTC 125

Qy       1659 TGTGCACAAGAGACCAGCTCCAGAAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGT 1718
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 TGTGCACAAGAGACCAGCTCCAGAAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGT 185

Qy       1719 CCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTGAGAGCA-GCACGCCGTCGT 1777
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        186 CCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTGAGAGCATGCACGCCGTCGT 245

Qy       1778 GGGCGAGCGATCACAATTACAATGCAGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGT 1837
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 GGGCGAGCGATCACAATTACAATGCAGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGT 305

Qy       1838 CACTGTTGTATAAATGTATGTATCACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTT 1897
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        306 CACTGTTGTATAAATGTATGTATCACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTT 365

Qy       1898 TCTGGATAGCCATCCCCTGGGCCTGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTG 1957
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 TCTGGATAGCCATCCCCTGGGCCTGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTG 425

Qy       1958 ATCACAGACACCGGCTGCACCATCAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTCCTG 2017
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 ATCACAGACACCGGCTGCACCATCAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTCCTG 485

Qy       2018 CTGCCCTGTGTCCATCCCTAGTCTGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTGTAA 2077
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 CTGCCCTGTGTCCATCCCTAGTCTGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTGTAA 545

Qy       2078 ACCTCACTGGTGAACGTTCACCTTAATGATTGATTCTTTAATCTCTGTTTTCACTCTCAG 2137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        546 ACCTCACTGGTGAACGTTCACCTTAATGATTGATTCTTTAATCTCTGTTTTCACTCTCAG 605

Qy       2138 GCTCTGGTAAGTATTTGTATTCTCTTCATCCCAGTCTGATTGCATAGCCACACTGCCCGG 2197
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        606 GCTCTGGTAAGTATTCGTATTCTCTTCATCCCAGTCTGATTGCATAGCCACACTGCCCGG 665

Qy       2198 CACGCCACATCCACCCCTGTCTGCACATGAGTTGTTCTGACAACAGCGCTGTATACGCTT 2257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 CACGCCACATCCACCCCTGTCTGCACATGAGTTGTTCTGACAACAGCGCTGTATACGCTT 725

Qy       2258 CAGTTTTTCCACATTGTCCACGGCCAGCACATGAAAGCATCACTTCTTTTTTATGTTGTG 2317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        726 CAGTTTTTCCACATTGTCCACGGCCAGCACATGAAAGCATCACTTCTTTTTTATGTTGTG 785

Qy       2318 GGAATCTTTGCAAGTTAGTGTTGCATCTGATTTTCAGGTGTACATTTATTTTTGACTGGG 2377
              ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db        786 GGAATCTTTGCAAGTKAGTGTTGCATCTGATTTTCAGGTGTACATTTATTTTTGACTGGG 845

Qy       2378 CAGATAGGGGATTTTTTTTTTTTTCCATGTCCGATTCACACGCTACACACCCACATGAAC 2437
              |||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db        846 CAGATAGGGGA--TTTTTTTTTTTCCATGTCCGATTCACACGCTACACACCCACATGAAC 903

Qy       2438 ACATTCGAACTTCGAAGG-CACACACTCCTGCTTCATAGGCCCCACGGTAAGTGAGTTCA 2496
              |||||||||||||||||| |||||||||||||||||||| ||||||||||||||||||||
Db        904 ACATTCGAACTTCGAAGGCCACACACTCCTGCTTCATAGGGCCCACGGTAAGTGAGTTCA 963

Qy       2497 CACCTAGAACACTGTCCTGACCGCAGGACGCGTGCCTTGGACTTGGT 2543
              ||||||||||||||:|||||||||| :||||||||||||||||||||
Db        964 CACCTAGAACACTGTYCTGACCGCAGAVCGCGTGCCTTGGACTTGGT 1010


RESULT 9
AL514138
LOCUS       AL514138               1201 bp    mRNA    linear   EST 08-MAY-2003
DEFINITION  AL514138 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
            CL0BB005ZB11 5-PRIME, mRNA sequence.
ACCESSION   AL514138
VERSION     AL514138.2  GI:30464023
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On Feb 13, 2001 this sequence version replaced gi:12777632.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
```

```
          Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
          Library was constructed by Life Technologies, a division of
          Invitrogen. This sequence belongs to sequence cluster 9799.r For
          more information about this cluster, see
          http://www.genoscope.cns.fr/
          cgi-bin/cluster.cgi?seq=CL0BB005ZB11RP1&cluster=9799.r. Contact :
          Feng Liang Email : fliang@lifetech.com URL :
          http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
          Faraday Avenue Genoscope sequence ID : CL0BB005ZB11RP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CL0BB005ZB11"
                     /tissue_type="NEUROBLASTOMA"
                     /clone_lib="Homo sapiens NEUROBLASTOMA"
                     /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN

  Query Match             33.3%;  Score 868;  DB 9;  Length 1201;
  Best Local Similarity   95.1%;  Pred. No. 1.3e-204;
  Matches  919;  Conservative    7;  Mismatches   37;  Indels    3;  Gaps    3;

Qy         65 GGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTG 124
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        160 GGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTGGTTGAAAATACGAAGCCTG 219

Qy        125 TTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCC 184
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        220 TAACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCC 279

Qy        185 AGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCAAG 244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        280 AGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCAAG 339

Qy        245 CTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTA 304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        340 CTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTA 399

Qy        305 AGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTA 364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 AGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTA 459

Qy        365 TCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCAC 424
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 TCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCAC 519

Qy        425 AGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCG 484
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 AGCAGCAGCCGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCG 579

Qy        485 TGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGG 544
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 TGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGG 639

Qy        545 AGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCA 604
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 AGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCA 699

Qy        605 GCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGG 664
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 GCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGG 759

Qy        665 AACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATA 724
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 AACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATA 819

Qy        725 GTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAAC 784
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        820 GTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAAC 879

Qy        785 AGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGG 844
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        880 AGGAGCCCACTGAGAGGCCCCTGAAA-GGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGG 938

Qy        845 AGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGC 904
              |||||||||||||||||||||||| |||||||||||||:|||||||:|||||||||||||
Db        939 AGGAGGGTCCCGCCGAGACTGTTGGCTCCGAGGCCAGTKACACTGTKGAGGGCGTCCTG- 997

Qy        905 CCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACA 964
              ||||||||||||||||||||||||||||||||||:| ||||||||||||||:||||||
Db        998 CCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTKT-CCAGGCTGGGAAARATGACA 1056

Qy        965 GAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACT 1024
              :||||||||: | |  |  | |||     | |   | |:    |||  ||| |
Db       1057 RAGAGAGTARTTGGAGGAAAGCGGCYAGGACACAAARAGAGGRSCTTGAGATTGGGGCGM 1116

Qy       1025 TGGGCC 1030
                | |||
Db       1117 CGAGCC 1122


RESULT 10
AL553387
LOCUS       AL553387               1201 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL553387 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI075YL03 5-PRIME, mRNA sequence.
ACCESSION   AL553387
VERSION     AL553387.2  GI:31275201
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12893174.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 9799.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI075CF02QP1&cluster=9799.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI075CF02QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI075YL03"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             33.1%;  Score 863.6;  DB 9;  Length 1201;
  Best Local Similarity   93.5%;  Pred. No. 1.6e-203;
  Matches  914;  Conservative   19;  Mismatches   33;  Indels   12;  Gaps    3;

Qy          1 CTCGGTGGCCGTCCGCCCACTCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGC 60
              | |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db         57 CCCGGGATCCGTCCGCCCACTCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGC 116

Qy         61 CTGCG------GAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAAT 114
              |||||      | :||||||||||||||||||||||||||||||||||||||||||||
Db        117 CTGCGGCCTGCGGASTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAAT 176

Qy        115 ACGAAGCCTGTTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCT 174
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        177 ACGAAGCCTGTTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCT 236

Qy        175 GCTTGGACCCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAA 234
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        237 GCTTGGACCCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAA 296

Qy        235 AGTGTCCAAGCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAG 294
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        297 AGTGTCCAAGCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAG 356

Qy        295 GAGGCACCTAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGA 354
              |||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db        357 GAGGCACCTAAGGCCATCAAACCCACCAGCAAARAGTTCAGGAAAACATGGGGTTTTCGA 416

Qy        355 AGGACCACTATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCG 414
              ||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||
Db        417 AGGACCACTATCGCCAAGCGARAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCG 476

Qy        415 CCACCCCCACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGC 474
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        477 CCACCCCCACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGC 536

Qy        475 ACTGAGCGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCT 534
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        537 ACTGAGCGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCT 596

Qy        535 GTCTCCCTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCC 594
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        597 GTCTCCCTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCC 656

Qy        595 TCCGAGGGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACA 654
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        657 TCCGAGGGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACA 716

Qy        655 GCTGTGAAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGAC 714
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        717 GCTGTGAAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGAC 776

Qy        715 ACCTCCGATAGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGG 774
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        777 ACCTCCGATAGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGG 836

Qy        775 AAGCGGGAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAG 834
              |||||||||||||||||||||||||||||||:|:|||||:||||||||||||||||||:
Db        837 AAGCGGGAACAGGAGCCCACTGAGAGGCCCCTGAAARGRATCCARAGTCGCCTGCGGAAR 896

Qy        835 AAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAG 894
              ||||||||||||||:|||||||||||||||||||||||||||||||| | |||||||:
Db        897 AAGCGCCGGGAGGARGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGA-AATGTGGAR 955

Qy        895 GGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGG 954
              |||||:|||||     |: | |::    :||    : |||| |:||||| :|   |:
Db        956 GGCGTYCTGCC-----ARTAAGMRGAGCCSAAAASATMAGGGGGTKTGTCCARGCTGGRA 1010

Qy        955 AAAGATGACAGAGAGAGT 972
              ||:||  | | ||  | |
Db       1011 AAWGAAAAAAAAGTAATT 1028


RESULT 11
AL550675/c
LOCUS       AL550675               1018 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL550675 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI056YK04 3-PRIME, mRNA sequence.
ACCESSION   AL550675
VERSION     AL550675.2  GI:31272492
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1018)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12887874.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 9799.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI056BF02NP1&cluster=9799.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI056BF02NP1.
FEATURES             Location/Qualifiers
     source          1. .1018
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI056YK04"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match           32.4%;  Score 845.2;  DB 9;  Length 1018;
  Best Local Similarity   97.5%;  Pred. No. 5.9e-199;
  Matches  856;  Conservative    9;  Mismatches   12;  Indels    1;  Gaps    1;

Qy       1189 AGGAATGGGGAAGACTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACT 1248
              | | |  |||:|||||||:||||||||||||||||||||||||||||||||||||||||
Db        879 AAGGAATGGGRAGACTATRTCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACT 820

Qy       1249 CATTCAGAAACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACC 1308
              |||||||| |||||||||||||||||||||||||||||||||||||||||||||||:||||
Db        819 CATTCAGANACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGRCACC 760

Qy       1309 GATTGTACAAGTATAGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGT 1368
              ||||||||||:||||||:||||||||||||||||||:|||||||||||||||||||||||
Db        759 GATTGTACAARTATAGGMACAATAGAGCAGAAGTCTMGCGAAGACCAAGGGATAAAGGGT 700

Qy       1369 AGAATTGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTG 1428
              :|||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        699 RGAATTGAGAAAGCTGCAAATCCAAGTGGCAAGAAGATACTCAAGATCTTCCAGCCTGTG 640

Qy       1429 ATAGAGGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCC 1488
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        639 ATAGAGGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCC 580

Qy       1489 GACTCGGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTT 1548
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        579 GACTCGGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTT 520

Qy       1549 CTAAGCTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAG 1608
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        519 CTAAGCTCAGGTAAAGNACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAG 460

Qy       1609 AAGCCCAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAG 1668
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        459 AAGCCCAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAG 400
```

```
Qy       1669 AGACCAGCTCCAGAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGG 1728
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 AGACCAGCTCCAGAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGG 340

Qy       1729 AGTGAAGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGAT 1788
             |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        339 AGTGAAGCACTCGGGAAGGAAGCTGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGAT 280

Qy       1789 CACAATTACAATGCAGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTAT 1848
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 CACAATTACAATGCAGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTAT 220

Qy       1849 AAATGTATGTATCACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCC 1908
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 AAATGTATGTATCACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCC 160

Qy       1909 ATCCCCTGGGCCTGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACAC 1968
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 ATCCCCTGGGCCTGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACAC 100

Qy       1969 CGGCTGCACCATCAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGT 2028
             ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db         99 CGGCTGCACCATCAGCGGGAAGCAGAGCCCATGTTCAGGATGCCTCCTGCTGCCCTGTGT 40

Qy       2029 CCATCCCTAGTCTGTCAGGACTTCCTGTCACTGTTTTC 2066
             ||||:|||||||||||| |||| |:||||| | |||||||
Db         39 CCATYCCTAGTCTGTNAGGA-TKCCTGTTATTGTTTTC 3
```

RESULT 12
BX404270
LOCUS BX404270 1050 bp mRNA linear EST 15-MAY-2003
DEFINITION BX404270 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone CL0BB022ZG10 5-PRIME, mRNA sequence.
ACCESSION BX404270
VERSION BX404270.1 GI:30764484
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1050)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 9799.r For more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CL0BB022ZG10RP1&cluster=9799.r. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope Sequence ID : CL0BB022ZG10RP1.
FEATURES Location/Qualifiers
source 1. .1050
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CL0BB022ZG10"
/tissue_type="NEUROBLASTOMA"
/clone_lib="Homo sapiens NEUROBLASTOMA"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."
ORIGIN

Query Match 32.1%; Score 838.2; DB 13; Length 1050;
Best Local Similarity 96.7%; Pred. No. 3.3e-197;
Matches 875; Conservative 12; Mismatches 10; Indels 8; Gaps 3;

```
Qy          2 TCGGTGGCCGTCCGCCCACTCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGCC 61
              ||   || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         52 TCCCGGGATGTCCGCCCACTCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGCC 111

Qy         62 TGCG------GAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATA 115
              ||||      | :||||||||||||||||||||||||||||||||||||||||||||||
Db        112 TGCGGCCTGCGGASTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATA 171

Qy        116 CGAAGCCTGTTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTG 175
              ||||||||||||||||||||||||||||||||||||:|||||||:||||||||||||||||
Db        172 CGAAGCCTGTTACTCGTGAACAGTGGCTGACAACAGTKTTGTTGTKAGCCTGGCTGTCTG 231

Qy        176 CTTGGACCCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAA 235
              ||||||||||||||||||||||||||||||||||||:|||:||||||||||||||||||||
Db        232 CTTGGACCCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTWTTTWGGATTTCAGGGAAAA 291

Qy        236 GTGTCCAAGCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGG 295
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        292 GTGTCCAAGCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCMATGAGG 351

Qy        296 AGGCACCTAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAA 355
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        352 AGGCACCTAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAA 411

Qy        356 GGACCACTATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGC 415
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        412 GGACCACTATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGC 471

Qy        416 CACCCCCACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCA 475
              ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db        472 CACCCCCACAGCAGCAGCTGGGCCTGTYCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCA 531

Qy        476 CTGAGCGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTG 535
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        532 CTGAGCGCGTTGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTG 591

Qy        536 TCTCCCTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCT 595
              |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        592 TYTCCCTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCT 651

Qy        596 CCGAGGGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAG 655
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        652 CCGAGGGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAG 711

Qy        656 CTGTGAAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACA 715
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        712 CTGTTAAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACA 771

Qy        716 CCTCCGATAGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGA 775
              ||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||
Db        772 CCTCCGATAGTGACAGCGATGGCCTGACCTTGAAAGAGCTTYAGAATCGCCTTCGCAGGA 831

Qy        776 AGCGGGAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGA 835
              ||||||||||||||||||||||||||||||:||| ||||||||||||||||||||||||||
Db        832 AGCGGGAACAGGAGCCCACTGAGAGGCCCCTSAAA-GGATCCAGAGTCGCCTGCGGAAGA 890

Qy        836 AGCGCCGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGG 895
              |||||||||||||||| |||||||||||||| ||||||||||||||:|||||||:||||
Db        891 AGCGCCGGGAGGAGGGT-CCGCCGAGACTGTTGGCTCCGAGGCCAGTKACACTGTKGAGG 949

Qy        896 GCGTC 900
              |||||
Db        950 GCGTC 954
```

RESULT 13
CD364822/c
LOCUS CD364822 850 bp mRNA linear EST 29-MAY-2003
DEFINITION UI-H-FT2-bjm-d-20-0-UI.s1 NCI_CGAP_FT2 Homo sapiens cDNA clone UI-H-FT2-bjm-d-20-0-UI 3', mRNA sequence.

```
ACCESSION   CD364822
VERSION     CD364822.1  GI:31148912
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 850)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Gary W. Hunninghake, U of I
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/cgap.html
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .850
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-H-FT2-bjm-d-20-0-UI"
                     /tissue_type="Aveolar Macrophage"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI_CGAP_FT2"
                     /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     NCI_CGAP_FT2 is a subtracted cDNA library constructed from
                     a pool of 81 RNA samples from Alveolar Macrophages
                     challenged with different treatments. The library was
                     subtracted according to Bonaldo, Lennon and Soares, Genome
                     Research, 6:791-806, 1996. The tissue was provided by Dr.
                     Gary W. Hunninghake of the University of Iowa.
                     TAG_TISSUE=Human Lung Aveolar Macrophage
                     TAG_LIB=UI-H-FT2
                     TAG_SEQ=GGCCATGCCG"
ORIGIN
```

Query Match 31.8%; Score 830.4; DB 14; Length 850;
Best Local Similarity 99.6%; Pred. No. 2.7e-195;
Matches 831; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1258 ACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACA 1317
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        850 ACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACA 791

Qy       1318 AGTATAGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAG 1377
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        790 AGTATAGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAG 731

Qy       1378 AAAGCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCG 1437
             ||||||||| |||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        730 AAAGCTGCANATCCAAGTGCNAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCG 671

Qy       1438 CCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTG 1497
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        670 CCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTG 611

Qy       1498 TACTGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCA 1557
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        610 TACTGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCA 551

Qy       1558 GGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGT 1617
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        550 GGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGT 491
```

```
Qy       1618 CTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCT 1677
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        490 CTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCT 431

Qy       1678 CCAGAAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCA 1737
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        430 CCAGAAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCA 371

Qy       1738 CTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTAC 1797
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        370 CTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTAC 311

Qy       1798 AATGCAGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATG 1857
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        310 AATGCAGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATG 251

Qy       1858 TATCACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGG 1917
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 TATCACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGG 191

Qy       1918 GCCTGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCAC 1977
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 GCCTGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCAC 131

Qy       1978 CATCAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTA 2037
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 CATCAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTA 71

Qy       2038 GTCTGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGTGAA 2091
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          70 GTCTGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGTGAA 17
```

```
RESULT 14
AK014422
LOCUS       AK014422                1614 bp    mRNA    linear   HTC 20-SEP-2003
DEFINITION  Mus musculus 18 days pregnant adult female placenta and extra
            embryonic tissue cDNA, RIKEN full-length enriched library,
            clone:3830408B01 product:death inducer-obliterator-1, full insert
            sequence.
ACCESSION   AK014422
VERSION     AK014422.1  GI:12852263
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
```

```
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 1614)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
            Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,
            Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
            Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
            Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
            Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
            Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
            Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K.,
            Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 10.0 and subtraction to
            Rot = 50.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
            with BamHI and XhoI. Vector: a modified pBluescript KS(+) after
            bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3'
            end: BamHI. Host: DH10B.
FEATURES             Location/Qualifiers
     source          1. .1614
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:3830408B01"
                     /db_xref="MGI:1910351"
                     /db_xref="taxon:10090"
                     /clone="3830408B01"
                     /sex="female"
                     /tissue_type="placenta and extra embryonic tissue"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="18 days pregnant adult"
     misc_feature    331. .>1614
                     /note="death inducer-obliterator-1 (MGD|MGI:1344352,
                     GB|AJ238332, evidence: BLASTN, 99%, match=2599)
                     putative"
ORIGIN

  Query Match           31.3%;  Score 817.8;  DB 11;  Length 1614;
  Best Local Similarity   73.1%;  Pred. No. 4.4e-192;
  Matches 1127;  Conservative    0;  Mismatches  387;  Indels   27;  Gaps    5;
```

```
Qy       21 TCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGCCTGCGGAGCTTACTCCACGG 80
Db      100 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 159

Qy       81 GAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTGTTACTCGTGAACAGTG 140
Db      160 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 219

Qy      141 GCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCCAGAGGTTTCGTCTGCC 200
Db      220 ACTGACAACAGTGGGGTGAG--GCTTGGCCGTCTGCTTGACC-----------TGGCCC 265

Qy      201 AGGGTTTTTGGTTGTATTTAGGATTT-CAGGGAAAAGTGTCCAAGCTTTCAGTGTTGGAG 259
Db      266 CAGGTCTATAATTTTATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGA 325

Qy      260 CAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTAAGGCCATCAAACCCA 319
Db      326 CAGGTATGGATGATAAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCA 385

Qy      320 CCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTATCGCCAAGCGAGAGG 379
Db      386 CCAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGG 445

Qy      380 GCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCACAGCAGCAGCTGGGCC 439
Db      446 GTGCAGGAGACACGGAGGTGGACCCCAGTGAGCAGCAACC------ACAGCAGCATAACC 499

Qy      440 TGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGA 499
Db      500 TCTCCCTGCGCCGCAGTGGACGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTA 559

Qy      500 CCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGC 559
Db      560 CCACGGTTCGGCGCCGAGGGAAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGC 619

Qy      560 CCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCAGCGTGGAAAGCGCTT 619
Db      620 CCACATCTTCCACAGTCACTGATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTT 679

Qy      620 CTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGGAACGACCAGCCTCTT 679
Db      680 CTGAGATCAGAAGTGCCCTGTA------TCTGACTCCTTAGGAAGAACATCCTGCCTCTT 733

Qy      680 CTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGGCC 739
Db      734 CTGAAAAGGCAAAAGGAGGTGAAGAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCT 793

Qy      740 TGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGA 799
Db      794 TTACGTTGAGGGAACTTCAGAACCGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGA 853

Qy      800 GGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGGTCCCGCCG 859
Db      854 GCTCCCTGAGAGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCG 913

Qy      860 AGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGC 919
Db      914 AAACTGGGAGTGTCCAAATAGGCAGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGC 973

Qy      920 CCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGG 979
Db      974 CTGAGGCTAGTCAGGGACCAGTGTCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGG 1033

Qy      980 AGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACTTGGGCCGACCGAAGC 1039
Db     1034 AAGGGAAGGCGACTCAGGGAAATACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGC 1093

Qy     1040 CTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACA 1099
Db     1094 CTGAGTGTGAGGTTTACGACCCCAATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACA 1153

Qy     1100 ACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCA 1159
```

```
Db       1154 ACAGGTTTATGATCTGCTGTGATCGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTA 1213

Qy       1160 TTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACT 1219

Db       1214 TTTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATT 1273

Qy       1220 GCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTA 1279

Db       1274 GCACCATTTTGCAAGTGCAGGATGAGACAAACGGTAGCGCCACCGATGAGCAGGACTCTG 1333

Qy       1280 AATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAGGAACAATAGAGCAGA 1339

Db       1334 GGTGCAGATCTGTGGGTGCTGATGGCACAGACTGCACAAGCATAGGGACAGTAGAGCAGA 1393

Qy       1340 AGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTGCAAATCCAAGTGGCA 1399

Db       1394 AGTCCGGAGAAGACCAGGGCATAAAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCA 1453

Qy       1400 AGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGGTGCCTCAAAATGTATTG 1459

Db       1454 AAAAAAAACTCAAGATATTCCAGCCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTG 1513

Qy       1460 GCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTGCAGTAATGACTGTATCC 1519

Db       1514 GCCCTGGGTGTTCCAGTGTAGCACAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTC 1573

Qy       1520 TCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGT 1560

Db       1574 TCAAACACGCAGCAGCTACCATGAGATTTCTAAGTTCAGGT 1614
```

```
RESULT 15
AL550431
LOCUS       AL550431                1119 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL550431 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI057YC01 5-PRIME, mRNA sequence.
ACCESSION   AL550431
VERSION     AL550431.2  GI:31272248
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1119)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12887402.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 9799.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI057AB01QP1&cluster=9799.r. Contact :
            Peng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI057AB01QP1.
FEATURES             Location/Qualifiers
     source          1. .1119
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI057YC01"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
```

ORIGIN

```
  Query Match           31.2%;  Score 813.4;  DB 9;  Length 1119;
  Best Local Similarity  92.5%;  Pred. No. 5e-191;
  Matches  856;  Conservative   18;  Mismatches   44;  Indels    7;  Gaps    2;

Qy         65 GGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTG 124

Db        157 GGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTG 216

Qy        125 TTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCC 184

Db        217 TWACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCC 276

Qy        185 AGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCAAG 244

Db        277 AGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCAAG 336

Qy        245 CTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTA 304

Db        337 CTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTA 396

Qy        305 AGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTA 364

Db        397 AGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTA 456

Qy        365 TCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCAC 424

Db        457 TCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCAC 516

Qy        425 AGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCG 484

Db        517 AGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCG 576

Qy        485 TGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGG 544

Db        577 TGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGG 636

Qy        545 AGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCA 604

Db        637 AGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCA 696

Qy        605 GCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGG 664

Db        697 GCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGG 756

Qy        665 AACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATA 724

Db        757 AACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATA 816

Qy        725 GTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAAC 784

Db        817 GTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAAC 876

Qy        785 AGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGG 844

Db        877 AGGAGCCCACTGAGAGGCCCCTGAAAGGAWTCCAGAGTCGCYTGCGGAAAAAGCGCCGGG 936

Qy        845 AGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGC 904

Db        937 AGGAGGGKTCCGCCGARACTKTGGGCCCGAGGGCARTK--AAATKTGGAAGGCGTCCTGC 994

Qy        905 CCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACA 964

Db        995 CAAKTAAGMAGARCCGAAAACATYA-----GGGGGTKTTTCCARGCTGGAARAWAAAAAA 1049

Qy        965 GAGAGAGTAAGTTGGAGGGAAAGGC 989

Db       1050 AATAATTKGAGGRAAAGSGSYAGGM 1074
```

Search completed: April 23, 2004, 18:02:49
Job time : 6463.74 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 23, 2004, 07:35:03 ; Search time 192.11 Seconds
(without alignments)
8281.924 Million cell updates/sec

Title: US-09-787-016A-2
Perfect score: 2867
Sequence: 1 tccgtggtagctctggaaat.....tttgtttccttttgtgctggg 2867

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	77.8	2.7	7218	1	US-08-232-463-14	Sequence 14, Appl
	2	55.4	1.9	6256	4	US-09-620-312D-269	Sequence 269, App
	3	40.2	1.4	7218	1	US-08-232-463-14	Sequence 14, Appl
c	4	37.8	1.3	505	4	US-09-621-976-15639	Sequence 15639, A
c	5	37.8	1.3	176373	3	US-09-128-155-17	Sequence 17, Appl
	6	37.2	1.3	6755	3	US-08-931-999-4	Sequence 4, Appli
	7	36.6	1.3	248	3	US-09-007-005-32	Sequence 32, Appl
	8	36.6	1.3	248	3	US-09-244-796-32	Sequence 32, Appl
	9	36.6	1.3	277	3	US-09-007-005-3	Sequence 3, Appli
	10	36.6	1.3	277	3	US-09-244-796-3	Sequence 3, Appli
	11	36.4	1.3	765	4	US-09-328-352-3047	Sequence 3047, Ap
c	12	36	1.3	474	4	US-09-621-976-18033	Sequence 18033, A
c	13	35.8	1.2	1492	3	US-09-196-520-7	Sequence 7, Appli
	14	35.6	1.2	289	3	US-09-007-005-17	Sequence 17, Appl
	15	35.6	1.2	289	3	US-09-244-796-17	Sequence 17, Appl
	16	35.6	1.2	2244	3	US-09-061-764A-18	Sequence 18, Appl
	17	35.6	1.2	2824	4	US-09-016-434-1349	Sequence 1349, Ap
	18	35	1.2	190	3	US-09-188-930-57	Sequence 57, Appl
	19	35	1.2	190	4	US-09-312-283C-57	Sequence 57, Appl
c	20	35	1.2	190	4	US-09-312-283C-402	Sequence 402, App
	21	35	1.2	472	4	US-09-621-976-18485	Sequence 18485, A
	22	34.8	1.2	1230025	4	US-09-198-452A-1	Sequence 1, Appli
c	23	34.6	1.2	245	1	US-08-139-937-8	Sequence 8, Appli
c	24	34.6	1.2	245	5	PCT-US93-11310-8	Sequence 8, Appli
	25	34.4	1.2	399	4	US-09-621-976-8976	Sequence 8976, Ap
c	26	34.2	1.2	1501	3	US-09-196-520-1	Sequence 1, Appli
c	27	34	1.2	732	4	US-09-134-000C-348	Sequence 348, App
c	28	34	1.2	1146	1	US-08-487-810-1	Sequence 1, Appli
	29	33.8	1.2	505	4	US-09-621-976-15639	Sequence 15639, A
	30	33.6	1.2	424	4	US-09-621-976-7975	Sequence 7975, Ap
	31	33.6	1.2	1886	4	US-09-594-506-31	Sequence 31, Appl
	32	33.6	1.2	2278	4	US-09-620-312D-1003	Sequence 1003, Ap
	33	33.6	1.2	2278	4	US-09-148-545-78	Sequence 78, Appl
	34	33.6	1.2	2278	4	US-09-148-545-131	Sequence 131, App
	35	33.4	1.2	1384	4	US-09-205-258-225	Sequence 225, App
c	36	33.4	1.2	5066	4	US-08-961-527-79	Sequence 79, Appl
c	37	33.4	1.2	5345	1	US-08-044-618-7	Sequence 7, Appli
	38	33.4	1.2	11049	4	US-10-204-708-23	Sequence 23, Appl
	39	33.4	1.2	11384	4	US-08-961-527-45	Sequence 45, Appl
c	40	33.4	1.2	202001	4	US-09-734-674-3	Sequence 3, Appli
	41	33.2	1.2	1497	4	US-09-489-039A-6080	Sequence 6080, Ap
c	42	33.2	1.2	1950	4	US-09-614-891-2	Sequence 2, Appli
c	43	33.2	1.2	2135	4	US-08-933-711B-17	Sequence 17, Appl
c	44	33.2	1.2	2501	4	US-09-614-891-3	Sequence 3, Appli
	45	33	1.2	364	4	US-09-621-976-17202	Sequence 17202, A

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, P.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 2.7%; Score 77.8; DB 1; Length 7218;

```
  Best Local Similarity   5.3%;  Pred. No. 6e-13;
  Matches   22;  Conservative  245;  Mismatches  152;  Indels    0;  Gaps    0;

Qy        602 TATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAAG 661
              |||||       |||  |||||       |  | |  | ::::: :::::::: ::::
Db       1474 TATCTATGCAAGTAGTTAAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRR 1415

Qy        662 AGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAACC 721
              :::::::::: :    :: :: :: :: :: ::    : :  :::::::    :::::
Db       1414 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1355

Qy        722 GCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAATC 781
              :     :::::::: ::::: :::::    : :::::::     :::::: ::  ::::
Db       1354 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1295

Qy        782 GCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGCA 841
              :   :::::::::: : ::::::::::    :  ::::  ::::: :   ::: ::: :
Db       1294 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1235

Qy        842 GTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTGT 901
              : :  ::: :::: :::      : ::: :::::   ::::  ::  :::::  :: :
Db       1234 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1175

Qy        902 CCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAATA 961
              ::  ::::: ::: :: : ::::::  ::  :::::::::::: ::   ::::::: :
Db       1174 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1115

Qy        962 CAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCC 1020
              :::::::::    ::::::: ::: :::  ::::   ::: : :::: | | | | |
Db       1114 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRATCGCAAGCTC 1056


RESULT 2
US-09-620-312D-269
; Sequence 269, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Chen, Rui-hong
;  APPLICANT:  Zhao, Qing A.
;  APPLICANT:  Wehrman, Tom
;  APPLICANT:  Xue, Aidong J.
;  APPLICANT:  Yang, Yonghong
;  APPLICANT:  Wang, Jian-Rui
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Ma, Yunqing
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Wang, Zhiwei
;  APPLICANT:  John Tillinghast
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. 6569662el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 784CIP2B
;  CURRENT APPLICATION NUMBER: US/09/620,312D
;  CURRENT FILING DATE:  2000-07-19
;  PRIOR APPLICATION NUMBER: 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 1105
;  SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 269
;   LENGTH: 6256
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (257)..(5443)
US-09-620-312D-269
```

```
  Query Match             1.9%;  Score 55.4;  DB 4;  Length 6256;
  Best Local Similarity   59.0%;  Pred. No. 4.7e-06;
  Matches   95;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy       1018 CCCAATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGT 1077
              ||||     | ||  |  | |||    | || ||   ||||||||||||  |  |||||
Db       1469 CCCAGCAAGCAGTGTGGGTTTTGCAAAAAACCACATGGCAACAGGTTTATGGTTGGCTGT 1528

Qy       1078 GATCGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGG 1137
              |   | ||||| || ||||| |||||||| ||||| ||  |    |    | | ||  | |
Db       1529 GGGAGATGTGATGACTGGTTTCATGGTGATTGTGTTGGGTTAAGTCTTTCTCAAGCACAG 1588

Qy       1138 CTCCTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTG 1178
              |   |||    | | |   |||| ||  ||||   ||| ||
Db       1589 CAGATGGGCGAGGAAGACAAAGAATATGTCTGTGTAAAATG 1629


RESULT 3
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match             1.4%;  Score 40.2;  DB 1;  Length 7218;
  Best Local Similarity   4.5%;  Pred. No. 0.27;
  Matches   18;  Conservative  209;  Mismatches  172;  Indels    0;  Gaps    0;

Qy       2465 TGTTTATAGTTTGAGTAAACTTTTGAAACCCAAGATTAGATCACTGAGTTTGGAGCCATT 2524
              ||  |  ||  | |||  | |||  ||  :::    ::   :: ::   :::    :: ::
```

```
Db       1040 TGGCTGCAGGTCGAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1099

Qy       2525 CTTGTTTCACTGCTGTACAAGGTACACCTTTCTTACTTTTTGTACAGTGGTTTCAAGTGA 2584

Db       1100 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1159

Qy       2585 ATTTTCATCTGTGTTCTCATGATAGTGTTTTGTTCCATCTCTGTCCACCCTCAGCCCCAG 2644

Db       1160 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1219

Qy       2645 ATAAGGGGGTCTCATGGCCTAAGCTGGCCTTAAATTTTCTGTAAAGTTGGGGGTCCTCCT 2704

Db       1220 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1279

Qy       2705 GCCCCCACCTGTAAAGAGTGCTGAAACTTACAGGCACACACCTGATGTATGCTGTGGGGT 2764

Db       1280 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1339

Qy       2765 GTGAACTCAGAGTGTTGGACCAGCAGTCTACCAGCTGAGCTGCAGTTCTAGCCATGGTAA 2824

Db       1340 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1399

Qy       2825 CTGAAACTGCTGATCTTCTTGTCTTTGTTTCCTTTGTGC 2863

Db       1400 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTAC 1438


RESULT 4
US-09-621-976-15639/c
; Sequence 15639, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 15639
;   LENGTH: 505
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-621-976-15639

  Query Match             1.3%;  Score 37.8;  DB 4;  Length 505;
  Best Local Similarity   13.5%;  Pred. No. 0.26;
  Matches   47;  Conservative  151;  Mismatches  148;  Indels    1;  Gaps    1;

Qy        641 AAAAGGCAAAAGGAGGTGAAGAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTA 700

Db        401 AMARRGWRCARGRTGSWGSWKGGYRMWGRGMAAAKMRRMAAAGGSYCGMTSYTSGSKMTG 342

Qy        701 CGTTGAAGGAACTTCAGAACCGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGT 760

Db        341 RKGSMTKRKRMMTYSGMWWTSYKCTKTGKKYTGWKSKKTRWTCTSWRKYMWMMSGCWARS 282

Qy        761 CCCTGAGAGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAAGATT-CTGCCGAA 819

Db        281 MKSWARSWYSMMACWCMMSASAYRARRSMYGARRSMMRAGAGWWRARRGKKRARGKSSMM 222

Qy        820 ACTGGGAGTGTCCAAATAGGCAGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCT 879

Db        221 RSKMSSMRMSAGKARMCRRWMWSCRRMSYSCMGSKCMSCRGTCAKMWRYARYAKRYASSM 162

Qy        880 GAGGCTAGTCAGGGACCAGTGTCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAA 939

Db        161 GKYMMGCRWCYAKCARMYGYYRSRSRSTGSRGMKYRRRKMYYMWKYWMWSWWCYRMGAAM 102

Qy        940 GGGAAGGCGACTCAGGGAAATACAGAGGAAAACCCCAGGGAAGCGGG 986


Db        101 YGMSARAYRMYASMSACKMCSRMMKMMSWSMWMRCWRSRYRCWMSGK 55


RESULT 5
US-09-128-155-17/c
; Sequence 17, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
;  APPLICANT: Pan, Yang
;  TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
;  TITLE OF INVENTION: AND USES THEREOF
;  FILE REFERENCE: 09404/052001
;  CURRENT APPLICATION NUMBER: US/09/128,155
;  CURRENT FILING DATE: 1998-08-03
;  EARLIER APPLICATION NUMBER: US 60/091,650
;  EARLIER FILING DATE: 1998-07-02
;  EARLIER APPLICATION NUMBER: US 60/054,646
;  EARLIER FILING DATE: 1997-08-04
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 17
;   LENGTH: 176373
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(176373)
;   OTHER INFORMATION: n = A,T,C or G
US-09-128-155-17

  Query Match             1.3%;  Score 37.8;  DB 3;  Length 176373;
  Best Local Similarity   51.5%;  Pred. No. 12;
  Matches   87;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy        482 AAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACTG 541

Db     115219 ACAAAAATTAGCAGGGCTTGGTGGTGTGTGCCTGTGGACCCAGCTACTCATGAGGCTAAG 115160

Qy        542 ATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCTG 601

Db     115159 GCAGGAAGAAAGCTTGAGCCTCAAGAGGTGGAGGCTGCAGTGAGCTGAGATCAAGCCACT 115100

Qy        602 TATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAA 650

Db     115099 GCACTCCAGCCTGAGCAACAGAACAAGATCCTGTCTCAAAAAAAATAAA 115051


RESULT 6
US-08-931-999-4
; Sequence 4, Application US/08931999
; Patent No. 6043219
;  GENERAL INFORMATION:
;    APPLICANT:  Iandolo, John J.
;    APPLICANT:  Crupper, Scott S.
;    TITLE OF INVENTION:  Broad Spectrum Chemotherapeutic Peptide
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hovey, Williams, Timmons & Collins
;      STREET:  2405 Grand Boulevard, Suite 400
;      CITY:  Kansas City
;      STATE:  Missouri
;      COUNTRY:  U.S.A.
;      ZIP:  64108
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/931,999
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US 08/710,561
;      FILING DATE:  19-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Collins, John M.
;      REGISTRATION NUMBER:  26,262
;      REFERENCE/DOCKET NUMBER:  25043-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  816/474-9050
;      TELEFAX:  816/474-9057
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6755 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Staphylococcus aureus
;      STRAIN:  UT0007
US-08-931-999-4

  Query Match             1.3%;  Score 37.2;  DB 3;  Length 6755;
  Best Local Similarity   46.5%;  Pred. No. 2.2;
  Matches  120;  Conservative    0;  Mismatches  138;  Indels    0;  Gaps    0;

Qy       1523 AACAAAAAACAAAACCCAAGGAAAAGGTCAAGACGAAGCCAGAAAAGTTCAGTCTTCCAA 1582
              || ||   | |||| | ||  ||||     ||  | ||| |  ||       ||| |   |||
Db       4466 AAAAAGGGAAAAAAACAAACAAAAAAAGGAAAGCAAAAACCAAAGCAGGAAGGCAAACAA 4525

Qy       1583 AATGCAGTGTTCAGGTGGGGATTAAAATCTCTTCTGTGCACAAGAGACTAGCGTCAGAGA 1642
                |   |      | |    |||  ||||       | |   | ||| |    |  | |||
Db       4526 CACAAAAGAGGCGGAGACGGAAGAAAAAACACCCCGAAAAAAACAACAAAAAGACAACAC 4585

Qy       1643 AAAGGGAAAACCCAGTGAAGAAAGTGATGCTGGCTTCCAGGAGTGAGACTTCTGGGAAGG 1702
              |||   ||  |  || ||||||||  |  |      | ||    |    |   |||
Db       4586 AAAACCAACCCAAAGAGAAGAAAGAAACACAACGAAAAAAGAAAACGGGGAAGGAAGAGG 4645

Qy       1703 AGGCAGCCTGTGAGAGCAGCACACCATCCTGGGCAAGTGACCACAACTACAATGCTGTGA 1762
              |  ||||| | || |    ||  |       || | | ||||| ||| | ||    | |
Db       4646 AAACAGCCAGAGAAACAGACAACCACAGAACGGGAGGGGACCAGAACAAAAAGAAGGCAA 4705

Qy       1763 AGCCAGAGAAGCCAGAGA 1780
              | |||||||||  | | |
Db       4706 AACCAGAGAAGGGACAAA 4723


RESULT 7
US-09-007-005-32
; Sequence 32, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350003
;  CURRENT APPLICATION NUMBER: US/09/007,005B
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32
;   LENGTH: 248
;   TYPE: RNA
;   ORGANISM: Homo sapiens
US-09-007-005-32

  Query Match             1.3%;  Score 36.6;  DB 3;  Length 248;
  Best Local Similarity   17.6%;  Pred. No. 0.38;
  Matches   42;  Conservative   91;  Mismatches  106;  Indels    0;  Gaps    0;

Qy        768 AGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAG 827
              :|: :|  : :|  :  ::|: : :|:     :|: : :|:  :  :|  :|:  |:|:
Db          1 RGRGRGRARCRARARURURARCRURARURURURARCRARARURURARCRARARURGRGRC 60

Qy        828 TGTCCAAATAGGCAGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAG 887
                |:|  :  |: :  |:  |: :|: :    :  ::: : :|:|:   :|:   :
Db         61 RURGRARARGRARARCRARGRARARARCRURGRARURCRURCRURGRARARGRARARGRA 120

Qy        888 TCAGGGACCAGTGTCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGAAGGC 947
               |: : :   :::  |  :: |:|:|  :   :|: : : :|  :   :|: : : : :
Db        121 RCRCRURGRCRURGRCRGRURARARARCRGRURCRGRURGRARARCRARGRCRURGRARA 180

Qy        948 GACTCAGGGAAATACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTG 1006
              :|   |: :|:| |  : :|: :|   |: :|: : :|:|:  :  |:|: :: :::::
Db        181 RARCRARCRARARARCRURGRGRARARCRARGRCRURGRCRGRURARARCRURCRURUR 239


RESULT 8
US-09-244-796-32
; Sequence 32, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350007
;  CURRENT APPLICATION NUMBER: US/09/244,796
;  CURRENT FILING DATE: 1999-02-05
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  EARLIER APPLICATION NUMBER: 09/007,005
;  EARLIER FILING DATE: 1998-01-14
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32
;   LENGTH: 248
;   TYPE: RNA
;   ORGANISM: Homo sapiens
US-09-244-796-32

  Query Match             1.3%;  Score 36.6;  DB 3;  Length 248;
  Best Local Similarity   17.6%;  Pred. No. 0.38;
  Matches   42;  Conservative   91;  Mismatches  106;  Indels    0;  Gaps    0;

Qy        768 AGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAG 827
              :|: :|  : :|  :  ::|: : :|:     :|: : :|:  :  :|  :|:  |:|:
Db          1 RGRGRGRARCRARARURURARCRURARURURURARCRARARURURARCRARARURGRGRC 60

Qy        828 TGTCCAAATAGGCAGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAG 887
                |:|  :  |: :  |:  |: :|: :    :  ::: : :|:|:   :|:   :
Db         61 RURGRARARGRARARCRARGRARARARCRURGRARURCRURCRURGRARARGRARARGRA 120

Qy        888 TCAGGGACCAGTGTCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGAAGGC 947
               |: : :   :::  |  :: |:|:|  :   :|: : : :|  :   :|: : : : :
Db        121 RCRCRURGRCRURGRCRGRURARARARCRGRURCRGRURGRARARCRARGRCRURGRARA 180

Qy        948 GACTCAGGGAAATACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTG 1006
              :|   |: :|:| |  : :|: :|   |: :|: : :|:|:  :  |:|: :: :::::
Db        181 RARCRARCRARARARCRURGRGRARARCRARGRCRURGRCRGRURARARCRURCRURUR 239


RESULT 9
US-09-007-005-3
; Sequence 3, Application US/09007005B
; Patent No. 6258558
```

```
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350003
;  CURRENT APPLICATION NUMBER: US/09/007,005B
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 277
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
US-09-007-005-3

  Query Match             1.3%;  Score 36.6;  DB 3;  Length 277;
  Best Local Similarity   17.6%;  Pred. No. 0.41;
  Matches   42;  Conservative   91;  Mismatches  106;  Indels    0;  Gaps    0;

Qy        768 AGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAG 827
              :|: :|  : :|  :  ::|: : :|:     :|: : :|:  :  :|  :|:  |:|:
Db          1 RGRGRGRARCRARARURURARCRURARURURURARCRARARURURARCRARARURGRGRC 60

Qy        828 TGTCCAAATAGGCAGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAG 887
                |:|  :  |: :  |:  |: :|: :    :  ::: : :|:|:   :|:   :
Db         61 RURGRARARGRARARCRARGRARARARCRURGRARURCRURCRURGRARARGRARARGRA 120

Qy        888 TCAGGGACCAGTGTCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGAAGGC 947
               |: : :   :::  |  :: |:|:|  :   :|: : : :|  :   :|: : : : :
Db        121 RCRCRURGRCRURGRCRGRURARARARCRGRURCRGRURGRARARCRARGRCRURGRARA 180

Qy        948 GACTCAGGGAAATACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTG 1006
              :|   |: :|:| |  : :|: :|   |: :|: : :|:|: :  |:|:| :: :::::
Db        181 RARCRARCRARARARCRURGRGRARARCRARGRCRURGRCRGRURARARCRURCRURUR 239


RESULT 10
US-09-244-796-3
; Sequence 3, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350007
;  CURRENT APPLICATION NUMBER: US/09/244,796
;  CURRENT FILING DATE: 1999-02-05
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  EARLIER APPLICATION NUMBER: 09/007,005
;  EARLIER FILING DATE: 1998-01-14
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 277
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
US-09-244-796-3

  Query Match             1.3%;  Score 36.6;  DB 3;  Length 277;
  Best Local Similarity   17.6%;  Pred. No. 0.41;
  Matches   42;  Conservative   91;  Mismatches  106;  Indels    0;  Gaps    0;

Qy        768 AGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAG 827
              :|: :|  : :|  :  ::|: : :|:     :|: : :|:  :  :|  :|:  |:|:
Db          1 RGRGRGRARCRARARURURARCRURARURURURARCRARARURURARCRARARURGRGRC 60

Qy        828 TGTCCAAATAGGCAGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAG 887
                |:|  :  |: :  |:  |: :|: :    :  ::: : :|:|:   :|:   :
Db         61 RURGRARARGRARARCRARGRARARARCRURGRARURCRURCRURGRARARGRARARGRA 120

Qy        888 TCAGGGACCAGTGTCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGAAGGC 947
               |: : :   :::  |  :: |:|:|  :   :|: : : :|  :   :|: : : : :
Db        121 RCRCRURGRCRURGRCRGRURARARARCRGRURCRGRURGRARARCRARGRCRURGRARA 180

Qy        948 GACTCAGGGAAATACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTG 1006
              :|   |: :|:| |  : :|: :|   |: :|: : :|:|: :  |:|:| :: :::::
Db        181 RARCRARCRARARARCRURGRGRARARCRARGRCRURGRCRGRURARARCRURCRURUR 239


RESULT 11
US-09-328-352-3047
; Sequence 3047, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3047
;   LENGTH: 765
;   TYPE: DNA
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-3047

  Query Match             1.3%;  Score 36.4;  DB 4;  Length 765;
  Best Local Similarity   50.6%;  Pred. No. 0.92;
  Matches   88;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy       2441 TAGATGCCCCAGATTTCAAATCAGTGTTTATAGTTTGAGTAAACTTTTGAAACCCAAGAT 2500
              ||||  ||||||||  ||||     |||  |     |||  |   |   |   || ||
Db        476 TAGAGTTACCAGATTTACAATCGCGACTTACACAAGCAGTGAGTGTAAAAGTGCCGAGCG 535

Qy       2501 TAGATCACTGAGTTTGGAGCCATTCTTGTTTCACTGCTGTACAAGGTACACCTTTCTTAC 2560
              |  |   |  | |    || || |||   ||||| |||||   | || || |    | |
Db        536 GAAGTTTATATGCTGATCGCTATGCTTTAGTCACTTCTGTAATGGCTAGACGTGGGATCC 595

Qy       2561 TTTTTGTACAGTGGTTTCAAGTGAATTTTCATCTGTGTTCTCATGATAGTGTTT 2614
              |||||  |||  | ||  ||   |||| | | |   |  | | ||   | ||
Db        596 ATTTTGATCAGCAGATTGTAGACTATTTACTTTTACATGGGCCTCATCAAGCTT 649


RESULT 12
US-09-621-976-18033/c
; Sequence 18033, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 18033
;   LENGTH: 474
;   TYPE: DNA
```

```
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 16
;   OTHER INFORMATION: n=a, g, c or t
US-09-621-976-18033

  Query Match             1.3%;  Score 36;  DB 4;  Length 474;
  Best Local Similarity   11.0%;  Pred. No. 0.9;
  Matches   24;  Conservative  107;  Mismatches   87;  Indels    0;  Gaps    0;

Qy       2292 TATTTCACATACTACACACCTACATAAACAGGTTTGAATTTTGAAGGTCATCTGCTGCTC 2351
Db        378 TWTTWAWCCYTTTTWWKRMMMYTTKRRRWWRGKWWTCYTAWKSCYWWYSRWKKKGRAAMY 319

Qy       2352 AGATCCCATGGTGAGTAATCGTGCTGACCATACACCTACAGCAGTATTCTCATCAAAGTG 2411
Db        318 KRRGGKYMRRAARRWTWARRAAMYTWWTYMWKGSCMWKSMMMRKWRKYMWKCCYWWWRKY 259

Qy       2412 AGAATTGTGGGTGGTAGCTTCCAGACTCGTAGATGCCCCAGATTTCAAATCAGTGTTTAT 2471
Db        258 CCMRKKRRRRGKKRARSYKGSYKGGSYTYYKGGKKGGGTKGGKAMSKKGRRRAMTTTYY 199

Qy       2472 AGTTTGAGTAAACTTTTGAAACCCAAGATTAGATCACT 2509
Db        198 KKKKWRSYWAARGKTTKKWAAMMMMMCMAWYMRSMMYY 161


RESULT 13
US-09-196-520-7/c
; Sequence 7, Application US/09196520
; Patent No. 6204039
; GENERAL INFORMATION:
;  APPLICANT: Falco, Carl S.
;  APPLICANT: Allen, Stephen M.
;  TITLE OF INVENTION: Plant Isocitrate Dehydrogenase Homologs
;  FILE REFERENCE: BB-1291
;  CURRENT APPLICATION NUMBER: US/09/196,520
;  CURRENT FILING DATE: 1998-11-19
;  EARLIER APPLICATION NUMBER: 60/067,388
;  EARLIER FILING DATE: December 2, 1997
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 1492
;   TYPE: DNA
;   ORGANISM: Triticum sp.
US-09-196-520-7

  Query Match             1.2%;  Score 35.8;  DB 3;  Length 1492;
  Best Local Similarity   46.9%;  Pred. No. 2.2;
  Matches  112;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy       1381 CCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCA 1440
Db        786 CTTGAGGAAAAGTCCGTCTGTTTTCCTCATGATATTAGCTTTATGTATCGCAGAGACCCT 727

Qy       1441 CAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATG 1500
Db        726 CTCCCTGCCATTGGCTTTGGCATAATGGAAAGCATACTCTGCCACTCTCAAACTTGCTTG 667

Qy       1501 AGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAACCCAAGGAAAAGGTCAAGACGAAG 1560
Db        666 GCGGGTAATAATTTTCAAACTTTCCACAACACCCCTCACAACCTGATGCTCAAGGCCACT 607

Qy       1561 CCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTGGGGATTAAAATCTCTTCTGT 1619
Db        606 ATACTCTCCTTCAGTATTTTCACGGATTGTCACAAGGTTCACATCATCATATCTAGTCT 548


RESULT 14
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350003
;  CURRENT APPLICATION NUMBER: US/09/007,005B
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

  Query Match             1.2%;  Score 35.6;  DB 3;  Length 289;
  Best Local Similarity   9.0%;  Pred. No. 0.86;
  Matches   24;  Conservative  104;  Mismatches  140;  Indels    0;  Gaps    0;

Qy        728 GGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAATCGCCTGA 787
Db          2 GRGRGRARCRARARURURARCRURARURURURARCRARARURURARCRARARURGRNRNR 61

Qy        788 GGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGCAGTGCCG 847
Db         62 SRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNR 121

Qy        848 AGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTGTCCCAGT 907
Db        122 SRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNR 181

Qy        908 CAGAGACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAATACAGAGG 967
Db        182 SRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRCRARGRCRURGRCRGRURARAR 241

Qy        968 AAAACCCCAGGGAAGCGGGCAAACCAAA 995
Db        242 CRURCRURURGRCRGRCRUAAAAAAAAA 269


RESULT 15
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350007
;  CURRENT APPLICATION NUMBER: US/09/244,796
;  CURRENT FILING DATE: 1999-02-05
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  EARLIER APPLICATION NUMBER: 09/007,005
;  EARLIER FILING DATE: 1998-01-14
;  NUMBER OF SEQ ID NOS: 33
```

```
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;    LENGTH: 289
;    TYPE: RNA
;    ORGANISM: Artificial Sequence
;    FEATURE:
;    OTHER INFORMATION: Translation template
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (1)...(289)
;    OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

  Query Match             1.2%;  Score 35.6;  DB 3;  Length 289;
  Best Local Similarity   9.0%;  Pred. No. 0.86;
  Matches   24;  Conservative  104;  Mismatches  140;  Indels    0;  Gaps    0;

Qy        728 GGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAATCGCCTGA 787
              |: : :|:|:|:|: : :|:| :: : : : :  |  :|: :  |:  |:|:: |    :
Db          2 GRGRGRARCRARARURURARCRURARURURURARCRARARURURARCRARARURGRNRNR 61

Qy        788 GGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGCAGTGCCG 847
              :: : : : : :::: : : : :  :    : :   ::: : :   : : : : ::    :
Db         62 SRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNR 121

Qy        848 AGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTGTCCCAGT 907
               : : :   :  :     :  :   : :  : : ::  :   : :  ::    :  :
Db        122 SRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNR 181

Qy        908 CAGAGACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAATACAGAGG 967
              :: : ::: : :     : : :   :   : : : : :   :  |: :|: : : : : :
Db        182 SRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRCRARGRCRURGRCRGRURARAR 241

Qy        968 AAAACCCCAGGGAAGCGGGCAAACCAAA 995
               : :|    :|: :|  :  |||  |||
Db        242 CRURCRURURGRCRGRCRUAAAAAAAAA 269


Search completed: April 23, 2004, 18:09:06
Job time : 197.11 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 23, 2004, 14:17:01 ; Search time 1363.62 Seconds
(without alignments)
9479.366 Million cell updates/sec

Title: US-09-787-016A-2
Perfect score: 2867
Sequence: 1 tccgtggtagctctggaaat.....tttgtttcctttgtgctggg 2867

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2907579 seqs, 2254313464 residues

Total number of hits satisfying chosen parameters: 5815158

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	558	19.5	1764	9	US-09-925-301-327	Sequence 327, App
c 2	354.6	12.4	645	15	US-10-029-386-25091	Sequence 25091, A
3	288	10.0	478	10	US-09-918-995-19127	Sequence 19127, A
c 4	117.8	4.1	536	15	US-10-029-386-11381	Sequence 11381, A
5	65	2.3	65	10	US-09-908-975-25239	Sequence 25239, A
6	57.8	2.0	915	9	US-09-764-864-471	Sequence 471, App
7	57.8	2.0	2320	15	US-10-175-523-142	Sequence 142, App
8	55.4	1.9	5975	10	US-09-873-367C-112	Sequence 112, App
9	55.4	1.9	6256	15	US-10-037-270-269	Sequence 269, App
10	55.4	1.9	6256	16	US-10-117-722-269	Sequence 269, App
11	55.4	1.9	6948	10	US-09-930-213-315	Sequence 315, App
12	50.4	1.8	653	15	US-10-184-644-402	Sequence 402, App
13	50.4	1.8	653	15	US-10-184-634-402	Sequence 402, App
14	48	1.7	469	13	US-10-085-783A-56748	Sequence 56748, A
15	48	1.7	469	16	US-10-242-535A-56748	Sequence 56748, A
16	48	1.7	483	9	US-09-864-761-2140	Sequence 2140, Ap
17	47.4	1.7	594	13	US-10-142-426-10	Sequence 10, Appl
18	47.4	1.7	594	15	US-10-123-155-10	Sequence 10, Appl
19	47.4	1.7	594	15	US-10-146-731-10	Sequence 10, Appl
20	47.4	1.7	594	15	US-10-140-472-10	Sequence 10, Appl
21	47.4	1.7	594	15	US-10-141-761-10	Sequence 10, Appl
22	47.4	1.7	594	15	US-10-142-885-10	Sequence 10, Appl
23	47.4	1.7	594	15	US-10-158-790-10	Sequence 10, Appl
24	47.4	1.7	594	16	US-10-137-871-10	Sequence 10, Appl
25	47.4	1.7	594	16	US-10-140-923-10	Sequence 10, Appl
26	47.4	1.7	594	16	US-10-141-756-10	Sequence 10, Appl
27	47.4	1.7	594	16	US-10-141-759-10	Sequence 10, Appl
28	47.4	1.7	594	16	US-10-140-805-10	Sequence 10, Appl
29	47.4	1.7	594	16	US-10-140-864-10	Sequence 10, Appl
30	42.8	1.5	114615	13	US-10-087-192-676	Sequence 676, App
31	41.4	1.4	745	13	US-10-142-426-68	Sequence 68, Appl
32	41.4	1.4	745	15	US-10-123-155-68	Sequence 68, Appl
33	41.4	1.4	745	15	US-10-146-731-68	Sequence 68, Appl
34	41.4	1.4	745	15	US-10-140-472-68	Sequence 68, Appl
35	41.4	1.4	745	15	US-10-141-761-68	Sequence 68, Appl
36	41.4	1.4	745	15	US-10-142-885-68	Sequence 68, Appl
37	41.4	1.4	745	15	US-10-158-790-68	Sequence 68, Appl
38	41.4	1.4	745	16	US-10-137-871-68	Sequence 68, Appl
39	41.4	1.4	745	16	US-10-140-923-68	Sequence 68, Appl
40	41.4	1.4	745	16	US-10-141-756-68	Sequence 68, Appl
41	41.4	1.4	745	16	US-10-141-759-68	Sequence 68, Appl
42	41.4	1.4	745	16	US-10-140-805-68	Sequence 68, Appl
43	41.4	1.4	745	16	US-10-140-864-68	Sequence 68, Appl
44	40.8	1.4	255	13	US-10-085-783A-34088	Sequence 34088, A
45	40.8	1.4	255	16	US-10-242-535A-34088	Sequence 34088, A

ALIGNMENTS

RESULT 1
US-09-925-301-327
; Sequence 327, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 327
; LENGTH: 1764
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1398)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1758)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1759)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1762)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-327

Query Match 19.5%; Score 558; DB 9; Length 1764;
Best Local Similarity 74.2%; Pred. No. 3.4e-166;

Matches 795; Conservative 0; Mismatches 250; Indels 27; Gaps 6;

```
Qy        984 GGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCCAATGCCCTGTACTGCATCTGCCG 1043
Db         32 GGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCCG 91

Qy       1044 CCAGCCTCACAACAACAGGTTTATGATCTGCTGTGATCGGTGTGAGGAGTGGTTCCATGG 1103
Db         92 CCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATGG 151

Qy       1104 TGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAACGGGGAAGACTA 1163
Db        152 CGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACTA 211

Qy       1164 CATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGATGAGACAAACGGTAGCGCCACCAA 1223
Db        212 TATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAGA 271

Qy       1224 TGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGATGGCACAGACTGCACAAGCATAGG 1283
Db        272 TCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAGG 331

Qy       1284 GACAGTAGAGCAGAAGTCCGGAGAAGACCAGGGCATAAAGGGTAGGATTGAGAAGGCAGC 1343
Db        332 AACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTGC 391

Qy       1344 AAACCCCAGCGGCAAGAAAAAACTCAAGATATTCCAGCCTGTCGTAGAGGCTCCTGGTGC 1403
Db        392 AAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGGTGC 451

Qy       1404 TCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCACAGCCTGACTCTGTGTATTGCAG 1463
Db        452 CTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCA-CCCGACTCGGTGTACTGCAG 510

Qy       1464 TAATGACTGCATTCTCAAACACGCAGCAGCTACCATGAGATTTCTAAGTTCAGGTAAAGA 1523
Db        511 TAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAAAGA 570

Qy       1524 ACAAAAAACAAAACCCAAGGAAAAGGTCAAGACGAAGCCAGAAAAGTTCAGTCTTCCAAA 1583
Db        571 ACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGAA 630

Qy       1584 ATGCAGTGTTCAGGTGGGGATTAAAATCTCTTCTGTGCACAAGAGACTAGCGTCAGAGAA 1643
Db        631 ATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAAA 690

Qy       1644 AAGGGAAAACCCAGTGAAGAA---AGTGATGCTGGCTTCCAGGAGTGAGACTTCTGGGAA 1700
Db        691 AAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGAA 750

Qy       1701 GGAGGCAGCCTGTGAGAGCAGCACACCATCCTGGGCAAGTGACCACAACTACAATGCTGT 1760
Db        751 GGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAGT 810

Qy       1761 GAAGCCAGAGAAGCCAGAGAAGCCCACTGCACTCTCGCCCACCCTATTGAGTAAATGTAC 1820
Db        811 AAAGCCAGAAAAGACTG---------CTGCTCCCTCGCCGTCACTGTTGTATAAATGTAT 861

Qy       1821 GTATCACCCAAAGGCTGGCTTCCCAGGCCCCTCCCATCATCT-----GGGTGGCTGCCTG 1875
Db        862 GTATCACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTG 921

Qy       1876 GGGCTGTCTAGGACCAG-----AGTCCTGGGTGTTCTGGTGCTGATAGTAGCCAGCAGC- 1929
Db        922 GGCCTGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCA 981

Qy       1930 ---TCACTGCCAGCCAGAAGCAGATACCAAGATGCCTCTGGACCCCAGGTGTTCCTGCCT 1986
Db        982 CCATCAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCT 1041

Qy       1987 AGCCTGTGGAGCCTCTCTGGGTGGTTCCTAAAGAGCTGTGTAGGCCTCATGT 2038
Db       1042 AGTCTGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGT 1093
```

```
RESULT 2
US-10-029-386-25091/c
; Sequence 25091, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GE
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 25091
;   LENGTH: 645
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL117379.14
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
;   OTHER INFORMATION: SWISSPROT HIT: P23583, EVALUE 2.50e-01
;   OTHER INFORMATION: EST_HUMAN HIT: AL580669.1, EVALUE 0.00e+00
;   OTHER INFORMATION: NT HIT: gi16193858, EVALUE 0.00e+00
US-10-029-386-25091
```

Query Match 12.4%; Score 354.6; DB 15; Length 645;
Best Local Similarity 76.4%; Pred. No. 1.3e-101;
Matches 464; Conservative 0; Mismatches 134; Indels 9; Gaps 2;

```
Qy        230 GTATGGATGATAAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCA 289
Db        645 GTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTAAGGCCATCAAACCCACCA 586

Qy        290 GTAAGGAGTTCAGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTG 349
Db        585 GCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTATCGCCAAGCGAGAGGGCG 526

Qy        350 CAGGAGACACGGAGGCGGACCCCAGTGAGCAGCAACC------ACAGCAGCATAACCTCT 403
Db        525 CAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCACAGCAGCAGCTGGGCCTGT 466

Qy        404 CCCTGCGCCGCAGTGGACGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCA 463
Db        465 CCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGACCA 406

Qy        464 CGGTTCGGCGCCGAGGGAAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCA 523
Db        405 TTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGCCCA 346

Qy        524 CATCTTCCACAGTCACTGATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTG 583
Db        345 CGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCAGCGTGGAAAGCGCTTCTG 286

Qy        584 AGATCAGAAGTGGCCCTGTATCTGACTCCTTA---GGGAAAGAACATCCTGCCTCTTCTG 640
Db        285 AGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGGAACGACCAGCCTCTTCTG 226

Qy        641 AAAAGGCAAAAGGAGGTGAAGAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTA 700
Db        225 AAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGGCCTGA 166

Qy        701 CGTTGAAGGAACTTCAGAACCGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGT 760
Db        165 CCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGAGGC 106

Qy        761 CCCTGAGAGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAA 820
Db        105 CCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCGAGA 46
```

```
Qy        821 CTGGGAG 827
              ||| | |
Db         45 CTGTGGG 39


RESULT 3
US-09-918-995-19127
; Sequence 19127, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 19127
;   LENGTH: 478
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(478)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-19127

  Query Match             10.0%;  Score 288;  DB 10;  Length 478;
  Best Local Similarity   77.7%;  Pred. No. 2e-80;
  Matches  348;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy        850 CAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTGTCCCAGTCA 909
              | | ||   |||  | ||||||||||||||  |||     ||||||    |||||||| |
Db         31 CGGCACGAGCCTGCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCT 90

Qy        910 GAGACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAATACAGAGGAA 969
              | || |||||||| ||| | | |||||||  || ||||||  ||||||   |  | ||| ||
Db         91 GGGAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAG 150

Qy        970 AACCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCCAATGCCCTG 1029
              | ||   | ||    ||||  ||| ||||||||  |||||||  ||||||||||||  ||||||
Db        151 GAGCCTGGAGACTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTG 210

Qy       1030 TACTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGATCGGTGTGAG 1089
              || ||||| ||||||||||||||||||||||||||||||||| ||||||||  || |||||
Db        211 TATTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAA 270

Qy       1090 GAGTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTCCTGGAAAGG 1149
              || ||||| ||||| || |||||||| ||||||||||| |||||| ||||  ||||||||
Db        271 GAATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGG 330

Qy       1150 AACGGGGAAGACTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGATGAGACAAAC 1209
              || |||||||||||| |||||||||| |||||||| ||||||||||||||||||||  |
Db        331 AATGGGGAAGACTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCAT 390

Qy       1210 GGTAGCGCCACCAATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGATGGCACAGAC 1269
                 |  || |||||||  ||   || ||| |||  | ||||||||||| ||
Db        391 TCAGAAACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGAT 450

Qy       1270 TGCACAAGCATAGGGACAGTAGAGCAGA 1297
              || ||||| ||||| ||| |||||||||
Db        451 TGTACAAGTATAGGAACAATAGAGCAGA 478


RESULT 4
US-10-029-386-11381/c
; Sequence 11381, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GE
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 11381
;   LENGTH: 536
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL117379.14
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
;   OTHER INFORMATION: SWISSPROT HIT: P80398, EVALUE 3.80e+00
;   OTHER INFORMATION: NT HIT: gi14769835, EVALUE 1.00e-78
;   OTHER INFORMATION: EST_HUMAN HIT: BF205331.1, EVALUE 2.00e-78
US-10-029-386-11381

  Query Match             4.1%;  Score 117.8;  DB 15;  Length 536;
  Best Local Similarity   83.2%;  Pred. No. 4.4e-26;
  Matches  134;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy        211 GCTTTCAGTTTTGGGACAGGTATGGATGATAAAGGGCACCTGAGCAATGAGGAAGCACCC 270
              |||| || |||   |  |||||||||| || |||||  ||| ||||||||||||| |||||
Db        167 GCTTACACTTTACCGCTAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCT 108

Qy        271 AAGGCTATCAAACCCACCAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCGAAGAACCACG 330
              ||||| |||||||||||||| || |||||||||||||  |||||||||||||| |||||
Db        107 AAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACT 48

Qy        331 ATTGCCAAACGTGAGGGTGCAGGAGACACGGAGGCGGACCC 371
              || ||||| || ||||| ||||| ||| ||||||| |||||
Db         47 ATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCC 7


RESULT 5
US-09-908-975-25239
; Sequence 25239, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
;  APPLICANT: SHOSHAN, Avi
;  APPLICANT:  WASSERMAN, Alon
;  APPLICANT:  MINTZ, Eli
;  APPLICANT:  MINTZ, Liat
;  APPLICANT:  FAIGLER, Simchon
;  TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE \
;  TITLE OF INVENTION:  THAT POPULATE A TRANSCRIPTOME
;  FILE REFERENCE: 36688-0005
;  CURRENT APPLICATION NUMBER: US/09/908,975
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: US 60/287,724
;  PRIOR FILING DATE: 2001-05-02
;  PRIOR APPLICATION NUMBER: US 60/221,607
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 32337
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 25239
;   LENGTH: 65
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-908-975-25239

  Query Match             2.3%;  Score 65;  DB 10;  Length 65;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-10;
  Matches   65;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2331 TTTGAAGGTCATCTGCTGCTCAGATCCCATGGTGAGTAATCGTGCTGACCATACACCTAC 2390
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTTGAAGGTCATCTGCTGCTCAGATCCCATGGTGAGTAATCGTGCTGACCATACACCTAC 60
```

```
Qy       2391 AGCAG 2395
              |||||
Db         61 AGCAG 65


RESULT 6
US-09-764-864-471
; Sequence 471, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 471
;   LENGTH: 915
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (894)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (903)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (905)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (914)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-471

  Query Match             2.0%;  Score 57.8;  DB 9;  Length 915;
  Best Local Similarity   66.4%;  Pred. No. 9.2e-07;
  Matches   83;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy       1001 AGTGTGAGGTTTACGACCCCAATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACA 1060
              |||  |||  |   ||         || | |||||||||||||||| | ||  ||| ||||
Db        281 AGTCCGAGAATGGGGAGAATGCGCCCATCTACTGCATCTGCCGCAAACCGGACATCAACT 340

Qy       1061 GGTTTATGATCTGCTGTGATCGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTT 1120
              | ||  ||||||  | |||||     ||   | ||||||||||||||| |||||  |   | ||
Db        341 GCTTCATGATCGGGTGTGACAACTGCAATGAGTGGTTCCATGGGGACTGCATCCGGATCA 400

Qy       1121 CTGAG 1125
              |||||
Db        401 CTGAG 405


RESULT 7
US-10-175-523-142
; Sequence 142, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
;  APPLICANT: Brockman, Jeffrey
;  APPLICANT:  Evans, David
;  APPLICANT:  Hook, Derek
;  APPLICANT:  Klimczak, Leszek
;  APPLICANT:  Laeng, Pascal
;  APPLICANT:  Palfreyman, Michael
;  APPLICANT:  Rajan, Prithi
;  TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
;  FILE REFERENCE: 3235/1J795-US3
;  CURRENT APPLICATION NUMBER: US/10/175,523
;  CURRENT FILING DATE:  2002-06-18
;  PRIOR APPLICATION NUMBER: US 60/299,151
;  PRIOR FILING DATE: 2001-06-18
;  PRIOR APPLICATION NUMBER: US 60/317,828
;  PRIOR FILING DATE: 2001-09-07
;  PRIOR APPLICATION NUMBER: US 60/325,150
;  PRIOR FILING DATE: 2001-09-25
;  PRIOR APPLICATION NUMBER: US 60/333,047
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: US 60/349,936
;  PRIOR FILING DATE: 2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/361,834
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 197
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 142
;   LENGTH: 2320
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-175-523-142

  Query Match             2.0%;  Score 57.8;  DB 15;  Length 2320;
  Best Local Similarity   66.4%;  Pred. No. 1.8e-06;
  Matches   83;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy       1001 AGTGTGAGGTTTACGACCCCAATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACA 1060
              |||  |||  |   ||         || | |||||||||||||||| | ||  ||| ||||
Db         109 AGTCCGAGAATGGGGAGAATGCGCCCATCTACTGCATCTGCCGCAAACCGGACATCAACT 168

Qy       1061 GGTTTATGATCTGCTGTGATCGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTT 1120
              | ||  ||||||  | |||||     ||   | ||||||||||||||| |||||  |   | ||
Db         169 GCTTCATGATCGGGTGTGACAACTGCAATGAGTGGTTCCATGGGGACTGCATCCGGATCA 228

Qy       1121 CTGAG 1125
              |||||
Db         229 CTGAG 233


RESULT 8
US-09-873-367C-112
; Sequence 112, Application US/09873367C
; Publication No. US20030165839A1
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  APPLICANT:  Soppet, Daniel
;  APPLICANT:  Endress, Gregory
;  APPLICANT:  Augustus, Meena
;  APPLICANT:  Ebner, Reinhard
;  APPLICANT:  Carter, Kenneth
;  TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
;  TITLE OF INVENTION:  Signature Gene Sets
;  FILE REFERENCE: 689290-64
;  CURRENT APPLICATION NUMBER: US/09/873,367C
;  CURRENT FILING DATE:  2003-04-29
;  PRIOR APPLICATION NUMBER: U.S. 60/236,891
;  PRIOR FILING DATE: 2000-09-29
;  PRIOR APPLICATION NUMBER: U.S. 60/236,842
;  PRIOR FILING DATE: 2000-09-29
;  PRIOR APPLICATION NUMBER: U.S. 60/244,867
;  PRIOR FILING DATE: 2000-11-01
;  PRIOR APPLICATION NUMBER: U.S. 60/245,084
;  PRIOR FILING DATE: 2000-11-01
;  NUMBER OF SEQ ID NOS: 1067
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 112
;   LENGTH: 5975
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-873-367C-112

  Query Match             1.9%;  Score 55.4;  DB 10;  Length 5975;
  Best Local Similarity   59.0%;  Pred. No. 2e-05;
  Matches   95;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy       1018 CCCAATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGT 1077
              ||||    | ||   |  | |||    | || ||   ||||||||||||||  |  |||||
Db       1200 CCCAGCAAGCAGTGTGGGTTTTGCAAAAAACCACATGGCAACAGGTTTATGGTTGGCTGT 1259
```

```
Qy       1078 GATCGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGG 1137
             |   | |||||  || ||||| |||||||| ||||| ||  |   |     | | ||  | |
Db       1260 GGGAGATGTGATGACTGGTTTCATGGTGATTGTGTTGGGTTAAGTCTTTCTCAAGCACAG 1319

Qy       1138 CTCCTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTG 1178
              |   |||    | | |   |||| ||  ||||   ||| ||
Db       1320 CAGATGGGCGAGGAAGACAAAGAATATGTCTGTGTAAAATG 1360


RESULT 9
US-10-037-270-269
; Sequence 269, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Chen, Rui-hong
;  APPLICANT:  Zhao, Qing A.
;  APPLICANT:  Wehrman, Tom
;  APPLICANT:  Xue, Aidong J.
;  APPLICANT:  Yang, Yonghong
;  APPLICANT:  Wang, Jian-Rui
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Ma, Yunqing
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Wang, Zhiwei
;  APPLICANT:  Tillinghast, John
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 784CIP2B
;  CURRENT APPLICATION NUMBER: US/10/037,270
;  CURRENT FILING DATE:  2002-01-04
;  PRIOR APPLICATION NUMBER: 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 1104
;  SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 269
;   LENGTH: 6256
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (257)..(5443)
US-10-037-270-269

  Query Match             1.9%;  Score 55.4;  DB 15;  Length 6256;
  Best Local Similarity   59.0%;  Pred. No. 2.1e-05;
  Matches   95;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy       1018 CCCAATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGT 1077
             ||||     | ||   |  | |||    | || ||  |||||||||||| | |||||
Db       1469 CCCAGCAAGCAGTGTGGGTTTTGCAAAAAACCACATGGCAACAGGTTTATGGTTGGCTGT 1528

Qy       1078 GATCGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGG 1137
             |   | |||||  || ||||| |||||||| ||||| ||  |   |     | | ||  | |
Db       1529 GGGAGATGTGATGACTGGTTTCATGGTGATTGTGTTGGGTTAAGTCTTTCTCAAGCACAG 1588

Qy       1138 CTCCTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTG 1178
              |   |||    | | |   |||| ||  ||||   ||| ||
Db       1589 CAGATGGGCGAGGAAGACAAAGAATATGTCTGTGTAAAATG 1629


RESULT 10
US-10-117-722-269
; Sequence 269, Application US/10117722
; Publication No. US20030219744A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. US20030219744A1el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 784CIP2BCIP
;  CURRENT APPLICATION NUMBER: US/10/117,722
;  CURRENT FILING DATE:  2002-04-04
;  PRIOR APPLICATION NUMBER: 09/620,312
;  PRIOR FILING DATE: 2000-07-19
;  PRIOR APPLICATION NUMBER: 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 1104
;  SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 269
;   LENGTH: 6256
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (257)..(5443)
US-10-117-722-269

  Query Match             1.9%;  Score 55.4;  DB 16;  Length 6256;
  Best Local Similarity   59.0%;  Pred. No. 2.1e-05;
  Matches   95;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy       1018 CCCAATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGT 1077
             ||||     | ||   |  | |||    | || ||  |||||||||||| | |||||
Db       1469 CCCAGCAAGCAGTGTGGGTTTTGCAAAAAACCACATGGCAACAGGTTTATGGTTGGCTGT 1528

Qy       1078 GATCGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGG 1137
             |   | |||||  || ||||| |||||||| ||||| ||  |   |     | | ||  | |
Db       1529 GGGAGATGTGATGACTGGTTTCATGGTGATTGTGTTGGGTTAAGTCTTTCTCAAGCACAG 1588

Qy       1138 CTCCTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTG 1178
              |   |||    | | |   |||| ||  ||||   ||| ||
Db       1589 CAGATGGGCGAGGAAGACAAAGAATATGTCTGTGTAAAATG 1629


RESULT 11
US-09-930-213-315
; Sequence 315, Application US/09930213
; Publication No. US20030170625A1
; GENERAL INFORMATION:
;  APPLICANT: ROSENTHAL, ANDRE
;  APPLICANT:  HINZMANN, BERND
;  APPLICANT:  SCHAFER, REINHARD
;  APPLICANT:  ZUBER, JOHANNES
;  APPLICANT:  TCHE-NITSE, OLEG
;  APPLICANT:  GRIPS, MARTIN
;  APPLICANT:  HELLNEGEL, MARTIN
;  APPLICANT:  SCHMITZ, ANNE-CHANTAL
;  APPLICANT:  SERS, CHRISTINE
;  TITLE OF INVENTION: DETECTION OF DIFFERENTIAL GENE EXPRESSIONS
;  FILE REFERENCE: ALBRE-14
;  CURRENT APPLICATION NUMBER: US/09/930,213
;  CURRENT FILING DATE:  2001-01-31
;  PRIOR APPLICATION NUMBER: DE 10004102.7
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 885
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 315
;   LENGTH: 6948
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-930-213-315
```

Query Match 1.9%; Score 55.4; DB 10; Length 6948;
Best Local Similarity 59.0%; Pred. No. 2.3e-05;
Matches 95; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```
Qy       1018 CCCAATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGT 1077
             ||||      | ||   |  | |||    | || ||   ||||||||||||| |  |||||
Db       2173 CCCAGCAAGCAGTGTGGGTTTTGCAAAAAACCACATGGCAACAGGTTTATGGTTGGCTGT 2232

Qy       1078 GATCGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGG 1137
              |   | ||||| || ||||| |||||||| ||||| ||  |    |  | | ||  | |
Db       2233 GGGAGATGTGATGACTGGTTTCATGGTGATTGTGTTGGGTTAAGTCTTTCTCAAGCACAG 2292

Qy       1138 CTCCTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTG 1178
              |  |||    | | |   |||| ||  ||||   ||| ||
Db       2293 CAGATGGGCGAGGAAGACAAAGAATATGTCTGTGTAAAATG 2333
```

RESULT 12
US-10-184-644-402
; Sequence 402, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 402
; LENGTH: 653
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-644-402

Query Match 1.8%; Score 50.4; DB 15; Length 653;
Best Local Similarity 6.6%; Pred. No. 0.00017;
Matches 43; Conservative 202; Mismatches 402; Indels 0; Gaps 0;

```
Qy        618 GAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAAGAGGAAGAAGACACCTC 677
              :::  : :  |: ::  :       :    :|:|    :     :    :    ::
Db          2 DRDLLRQSLNCHGSSLLSLLRSEQQDNPHFRSLLGSAAEPARGPPPQHPLQGRKEKRVDN 61

Qy        678 TGACAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAACCGCCTTCGGAGAAAGCG 737
                  :  |:     :   :| |     :   ::          :   : :::    :
Db         62 IEIQKFISKKADLLFALSWKSDAPATSEINEDSEDHYAIMPPLEQFMEIPSMDRRELFFR 121

Qy        738 AGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAATCGCCTGAGGAAGAAGCG 797
              :  : : :     : :  |     | | | :   :     | |   :: :     : :
Db        122 DIERGDIVIGRISSIREFGFFMVLICLGSGIMRDIAHLEITALCPLRDVPSHSNHGDPLS 181

Qy        798 CAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGCAGTGCCGAGCAGGACAG 857
              :   |: :|| :: ::::   |:   : :   :     : |:  :         :::   :
Db        182 YYQTGDIIRAGIKDIDRYHEKLAVSLYSSSLPPHLSGIKLGVISSEELPLYYRRSVELNS 241

Qy        858 ACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTGTCCCAGTCAGAGACAGA 917
              :  ::  :  :   :  :       : |  :   : : ::   :   :  : |:| ::
Db        242 NSLESYENVMQSSLGFVNPGVVEFLLEKLGIDESNPPSLMRGLQSKNPSEDDFASALRKK 301

Qy        918 TGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAATACAGAGGAAAACCCCAG 977
              :|::  :  :   ::  ::|:::: :    :|  ::  :  | : :||  |   :   |
Db        302 QSASWALKCVKIGVDYFKVGRHVDAMNEYNKALEIDKQNVEALVARGALYATKGSLNKAI 361
```

```
Qy        978 GGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCCAATGCCCTGTACTGCAT 1037
              :       | : |:  |  : : |             ::  :   : | :  :
Db        362 EDFELALENCPTHRNARKYLCQTLVERGGQLEEEEKFLNAESYYKKALALDETFKDAEDA 421

Qy        1038 CTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGATCGGTGTGAGGAGTGGTT 1097
                : :::: ::   :        : :|:     :  :     :: :::::  ::
Db         422 LQKLHKYMQKSLELREKQAEKEEKQKTKKIETSAEKLRKLLKEEKRLKKKRRKSTSSSSV 481

Qy        1098 CCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAACGGGGA 1157
               ::|: : : : : : :| :::::::: :: ::: ::::::  ::   :  ::: | : |
Db         482 SSADESVSSSSSSSSSGHKRHKKHKRNRSESSRSSRRHSSRASSNQIDQNRKDECYPVPA 541

Qy        1158 AGACTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGATGAGACAAACGGTAGCGC 1217
                    :   :     : :  : :|  :  ::    ::     :  :::  :
Db         542 NTSASFLNHKQEVEKLLGKQDRLQYEKTQIKEKDRCPLSSSSLEIPDDFGVYSYLFKKLT 601

Qy        1218 CACCAATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGATGGCA 1264
                  |   |     :  :: : :: :  :  :| :: : ::| :
Db         602 IKQPQAGPSGDIPEEGIVIIDDSSIHVTDPEDLQVGQDMEVEDSGID 648
```

RESULT 13
US-10-184-634-402
; Sequence 402, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 402
; LENGTH: 653
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-634-402

Query Match 1.8%; Score 50.4; DB 15; Length 653;
Best Local Similarity 6.6%; Pred. No. 0.00017;
Matches 43; Conservative 202; Mismatches 402; Indels 0; Gaps 0;

```
Qy        618 GAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAAGAGGAAGAAGACACCTC 677
              :::  : :  |: ::  :       :    :|:|    :     :    :    ::
Db          2 DRDLLRQSLNCHGSSLLSLLRSEQQDNPHFRSLLGSAAEPARGPPPQHPLQGRKEKRVDN 61

Qy        678 TGACAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAACCGCCTTCGGAGAAAGCG 737
                  :  |:     :   :| |     :   ::          :   : :::    :
Db         62 IEIQKFISKKADLLFALSWKSDAPATSEINEDSEDHYAIMPPLEQFMEIPSMDRRELFFR 121

Qy        738 AGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAATCGCCTGAGGAAGAAGCG 797
              :  : : :     : :  |     | | | :   :     | |   :: :     : :
Db        122 DIERGDIVIGRISSIREFGFFMVLICLGSGIMRDIAHLEITALCPLRDVPSHSNHGDPLS 181

Qy        798 CAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGCAGTGCCGAGCAGGACAG 857
              :   |: :|| :: ::::   |:   : :   :     : |:  :         :::   :
Db        182 YYQTGDIIRAGIKDIDRYHEKLAVSLYSSSLPPHLSGIKLGVISSEELPLYYRRSVELNS 241

Qy        858 ACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTGTCCCAGTCAGAGACAGA 917
              :  ::  :  :   :  :       : |  :   : : ::   :   :  : |:| ::
```

```
Db        242 NSLESYENVMQSSLGFVNPGVVEFLLEKLGIDESNPPSLMRGLQSKNFSEDDFASALRKK 301

Qy        918 TGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAATACAGAGGAAAACCCCAG 977

Db        302 QSASWALKCVKIGVDYFKVGRHVDAMNEYNKALEIDKQNVEALVARGALYATKGSLNKAI 361

Qy        978 GGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCCAATGCCCTGTACTGCAT 1037

Db        362 EDFELALENCPTHRNARKYLCQTLVERGGQLEEEEKFLNAESYYKKALALDETFKDAEDA 421

Qy       1038 CTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGATCGGTGTGAGGAGTGGTT 1097

Db        422 LQKLHKYMQKSLELREKQAEKEEKQKTKKIETSAEKLRKLLKEEKRLKKKRRKSTSSSSV 481

Qy       1098 CCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAACGGGGA 1157

Db        482 SSADESVSSSSSSSSSGHKRHKKHKRNRSESSRSSRRHSSRASSNQIDQNRKDECYPVPA 541

Qy       1158 AGACTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGATGAGACAAACGGTAGCGC 1217

Db        542 NTSASFLNHKQEVEKLLGKQDRLQYEKTQIKEKDRCPLSSSSLEIPDDFGVYSYLFKKLT 601

Qy       1218 CACCAATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGATGGCA 1264

Db        602 IKQPQAGPSGDIPEEGIVIIDDSSIHVTDPEDLQVGQDMEVEDSGID 648


RESULT 14
US-10-085-783A-56748
; Sequence 56748, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
;  APPLICANT: ChondroGene Inc.
;  APPLICANT:  Liew, C.C.
;  TITLE OF INVENTION: Compositions and Methods Relatiing to Osteoarthritis
;  FILE REFERENCE: 4231/2002
;  CURRENT APPLICATION NUMBER: US/10/085,783A
;  CURRENT FILING DATE:  2002-02-28
;  PRIOR APPLICATION NUMBER: US 60/305,340
;  PRIOR FILING DATE: 2001-07-13
;  PRIOR APPLICATION NUMBER: US 60/275,017
;  PRIOR FILING DATE: 2001-03-12
;  PRIOR APPLICATION NUMBER: US 60/271,955
;  PRIOR FILING DATE: 2001-02-28
;  NUMBER OF SEQ ID NOS: 58994
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 56748
;   LENGTH: 469
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (5)..(5)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (30)..(30)
;   OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-56748

  Query Match             1.7%;  Score 48;  DB 13;  Length 469;
  Best Local Similarity   62.5%;  Pred. No. 0.00076;
  Matches   75;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy       1059 CAGGTTTATGATCTGCTGTGATCGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTAT 1118

Db        116 CAGGTTTATGGTTGGCTGTGGGAGATGTGATGACTGGTTTCATGGTGATTGTGTTGGGTT 175

Qy       1119 TTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTG 1178

Db        176 AAGTCTTTCTCAAGCACAGCAGATGGGCGAGGAAGACAAAGAATATGTCTGTGTAAAATG 235


RESULT 15
US-10-242-535A-56748
; Sequence 56748, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
;  APPLICANT: ChondroGene Inc.
;  APPLICANT:  Liew, C.C.
;  TITLE OF INVENTION: Compositions and Methods Relatiing to Osteoarthritis
;  FILE REFERENCE: 4231/2005
;  CURRENT APPLICATION NUMBER: US/10/242,535A
;  CURRENT FILING DATE:  2002-09-12
;  PRIOR APPLICATION NUMBER: US 10/085,783
;  PRIOR FILING DATE: 2002-02-28
;  PRIOR APPLICATION NUMBER: US 60/305,340
;  PRIOR FILING DATE: 2001-07-13
;  PRIOR APPLICATION NUMBER: US 60/275,017
;  PRIOR FILING DATE: 2001-03-12
;  PRIOR APPLICATION NUMBER: US 60/271,955
;  PRIOR FILING DATE: 2001-02-28
;  NUMBER OF SEQ ID NOS: 58994
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 56748
;   LENGTH: 469
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (5)..(5)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (30)..(30)
;   OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-56748

  Query Match             1.7%;  Score 48;  DB 16;  Length 469;
  Best Local Similarity   62.5%;  Pred. No. 0.00076;
  Matches   75;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy       1059 CAGGTTTATGATCTGCTGTGATCGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTAT 1118

Db        116 CAGGTTTATGGTTGGCTGTGGGAGATGTGATGACTGGTTTCATGGTGATTGTGTTGGGTT 175

Qy       1119 TTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTG 1178

Db        176 AAGTCTTTCTCAAGCACAGCAGATGGGCGAGGAAGACAAAGAATATGTCTGTGTAAAATG 235


Search completed: April 23, 2004, 23:50:40
Job time : 1366.62 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 23, 2004, 04:38:03 ; Search time 961.654 Seconds
(without alignments)
11529.921 Million cell updates/sec

Title: US-09-787-016A-1
Perfect score: 2610
Sequence: 1 ctcggtggccgtccgcccac.....actcttaagatcatatcctg 2610

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2610	100.0	2610	3	AAZ90578	Aaz90578 Human dea
	2	2516.2	96.4	2772	4	AAH15102	Aah15102 Human cDN
	3	1810.4	69.4	2085	5	ABV28029	Abv28029 Human pro
	4	1810.4	69.4	2085	5	ABV27935	Abv27935 Human pro
	5	1810.4	69.4	2085	5	ABV22095	Abv22095 Human pro
	6	1808.8	69.3	2085	5	ABV22192	Abv22192 Human pro
	7	1564.4	59.9	1764	3	AAC77933	Aac77933 Human can
	8	1280.6	49.1	7290	4	AAK53352	Aak53352 Human pol
	9	1280.6	49.1	7838	4	AAK52368	Aak52368 Human pol
	10	1190.8	45.6	2332	5	AAS86576	Aas86576 DNA encod
	11	1023.6	39.2	2867	3	AAZ90579	Aaz90579 Murine de
	12	948.4	36.3	30626	4	AAK67051	Aak67051 Human imm
	13	682.6	26.2	2986	5	AAS86575	Aas86575 DNA encod
	14	596.8	22.9	659	3	AAC08973	Aac08973 Human sec
	15	510.8	19.6	681	4	AAH08005	Aah08005 Human cDN
	16	503	19.3	503	5	AAS86574	Aas86574 DNA encod
	17	435	16.7	478	8	ACH31915	Ach31915 Human end
c	18	414.4	15.9	481	5	ABV42461	Abv42461 Human pro
c	19	336.6	12.9	542	4	AAH12970	Aah12970 Human cDN
	20	314.2	12.0	362	5	ABV12528	Abv12528 Human pro
	21	305.6	11.7	365	5	ABV42575	Abv42575 Human pro
	22	305.6	11.7	365	5	ABV33664	Abv33664 Human pro
	23	249.8	9.6	256	5	ABV03359	Abv03359 Human pro
	24	235.2	9.0	408	3	AAA43780	Aaa43780 Mouse sec
	25	231.2	8.9	324	6	ABN21248	Abn21248 Human ORF
c	26	209	8.0	7838	4	AAK52368	Aak52368 Human pol
	27	202.4	7.8	2310	3	AAC74383	Aac74383 Human sec
	28	182.6	7.0	7174	9	ADC37602	Adc37602 Human nuc
c	29	137.2	5.3	636	5	ABV03229	Abv03229 Human pro
c	30	102.6	3.9	593	5	ABV12398	Abv12398 Human pro
c	31	101.2	3.9	179	5	ABV33540	Abv33540 Human pro
c	32	91.2	3.5	226	6	ABZ08454	Abz08454 Human leu
	33	78.2	3.0	6051	4	ABL29341	Abl29341 Drosophil
	34	68.6	2.6	8751	4	ABL29340	Abl29340 Drosophil
	35	65	2.5	5975	6	ABL61775	Abl61775 Colon ade
	36	65	2.5	6256	4	AAI58390	Aai58390 Human pol
	37	65	2.5	6256	8	ADB48359	Adb48359 Novel hum
	38	65	2.5	6584	4	AAI60176	Aai60176 Human pol
	39	65	2.5	6947	7	ACF34533	Acf34533 Gene enco
	40	65	2.5	6948	5	AAH81806	Aah81806 Human dif
	41	65	2.5	6996	4	AAK53025	Aak53025 Human pol
	42	65	2.5	7354	4	AAK52041	Aak52041 Human pol
	43	60	2.3	60	6	ABN33521	Abn33521 Human spl
	44	57.4	2.2	2000	7	ADA71938	Ada71938 Rice gene
	45	56.8	2.2	10732	3	AAA10594	Aaa10594 Gene enco

ALIGNMENTS

RESULT 1
AAZ90578
ID AAZ90578 standard; cDNA; 2610 BP.
XX
AC AAZ90578;
XX
DT 19-JUN-2000 (first entry)
XX
DE Human death inducer-obliterator 1 (DIO-1) polypeptide encoding cDNA.
XX
KW Death inducer-obliterator 1; DIO-1; cell death; cancer; tumour; human;
KW autoimmune disease; cytostatic; immunosuppressive; antidiabetic;
KW antirheumatic; antiinflammatory; antiproliferative; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 265. .1953
FT /*tag= a
FT /product= "DIO-1"
XX
PN WO200015787-A1.
XX
PD 23-MAR-2000.
XX
PF 10-SEP-1999; 99WO-GB003019.
XX
PR 10-SEP-1998; 98SE-00003069.
PR 17-SEP-1998; 98US-0100873P.
XX
PA (CNSJ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.
PA (BANN/) BANNERMAN D G.
XX
PI Alonso CM, Domingo DG, Grandien A, Leonardo E, Martinez P;
XX
DR WPI; 2000-271426/23.
DR P-PSDB; AAY67579.
XX
PT New DNA encoding human and murine death inducer-obliterator 1
PT polypeptides, useful in the treatment of cancer, autoimmune diseases,
PT diabetes, rheumatoid arthritis, benign tumors, malignant tumors and
PT hyperproliferative skin disorders.
XX
PS Claim 1; Fig 1A; 27pp; English.
XX
CC The invention provides nucleic acids encoding the human and murine death

CC inducer-obliterator 1 (DIO-1) polypeptides. The polypeptides can be
CC expressed by standard recombinant methodology. The DIO-1 polypeptides,
CC agonists and antagonists are used as a medicament for treating diseases
CC characterized by an alteration in cell death or by hyperproliferation,
CC e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign
CC tumours, malignant tumours or hyperproliferative skin disorders. They are
CC also useful in the treatment of metabolic, proliferative or inflammatory
CC conditions. The present sequence represents a cDNA encoding the human DIO
CC -1 polypeptide
XX
SQ Sequence 2610 BP; 644 A; 667 C; 734 G; 565 T; 0 U; 0 Other;

Query Match 100.0%; Score 2610; DB 3; Length 2610;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2610; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 CTCGGTGGCCGTCCGCCCACTCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGC 60
Db    1 CTCGGTGGCCGTCCGCCCACTCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGC 60

Qy   61 CTGCGGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAG 120
Db   61 CTGCGGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAG 120

Qy  121 CCTGTTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGG 180
Db  121 CCTGTTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGG 180

Qy  181 ACCCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTC 240
Db  181 ACCCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTC 240

Qy  241 CAAGCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCA 300
Db  241 CAAGCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCA 300

Qy  301 CCTAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACC 360
Db  301 CCTAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACC 360

Qy  361 ACTATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCC 420
Db  361 ACTATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCC 420

Qy  421 CCACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAG 480
Db  421 CCACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAG 480

Qy  481 CGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCC 540
Db  481 CGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCC 540

Qy  541 CTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAG 600
Db  541 CTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAG 600

Qy  601 GGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTG 660
Db  601 GGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTG 660

Qy  661 AAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCC 720
Db  661 AAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCC 720

Qy  721 GATAGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGG 780
Db  721 GATAGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGG 780

Qy  781 GAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGC 840
Db  781 GAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGC 840

Qy  841 CGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTC 900
Db  841 CGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTC 900

Qy  901 CTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGAT 960
Db  901 CTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGAT 960

Qy  961 GACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGA 1020
Db  961 GACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGA 1020

Qy 1021 GACTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATT 1080
Db 1021 GACTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATT 1080

Qy 1081 TGCCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTT 1140
Db 1081 TGCCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTT 1140

Qy 1141 CATGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAA 1200
Db 1141 CATGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAA 1200

Qy 1201 GACTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACG 1260
Db 1201 GACTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACG 1260

Qy 1261 GCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGT 1320
Db 1261 GCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGT 1320

Qy 1321 ATAGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAA 1380
Db 1321 ATAGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAA 1380

Qy 1381 GCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCT 1440
Db 1381 GCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCT 1440

Qy 1441 GGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTAC 1500
Db 1441 GGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTAC 1500

Qy 1501 TGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGT 1560
Db 1501 TGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGT 1560

Qy 1561 AAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTT 1620
Db 1561 AAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTT 1620

Qy 1621 CCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCA 1680
Db 1621 CCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCA 1680

Qy 1681 GAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTC 1740
Db 1681 GAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTC 1740

Qy 1741 GGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAAT 1800
Db 1741 GGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAAT 1800

Qy 1801 GCAGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTAT 1860
Db 1801 GCAGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTAT 1860

Qy 1861 CACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCC 1920
Db 1861 CACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCC 1920

Qy 1921 TGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCAT 1980
```

```
Db       1921 TGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCAT 1980

Qy       1981 CAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 CAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTC 2040

Qy       2041 TGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGTGAACGTTCACCT 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGTGAACGTTCACCT 2100

Qy       2101 TAATGATTGATTCTTTAATCTCTGTTTTCACTCTCAGGCTCTGGTAAGTATTTGTATTCT 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TAATGATTGATTCTTTAATCTCTGTTTTCACTCTCAGGCTCTGGTAAGTATTTGTATTCT 2160

Qy       2161 CTTCATCCCAGTCTGATTGCATAGCCACACTGCCCGGCACGCCACATCCACCCCTGTCTG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 CTTCATCCCAGTCTGATTGCATAGCCACACTGCCCGGCACGCCACATCCACCCCTGTCTG 2220

Qy       2221 CACATGAGTTGTTCTGACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CACATGAGTTGTTCTGACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGG 2280

Qy       2281 CCAGCACATGAAAGCATCACTTCTTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTG 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCAGCACATGAAAGCATCACTTCTTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTG 2340

Qy       2341 CATCTGATTTTCAGGTGTACATTTATTTTTGACTGGGCAGATAGGGGATTTTTTTTTTTT 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CATCTGATTTTCAGGTGTACATTTATTTTTGACTGGGCAGATAGGGGATTTTTTTTTTTT 2400

Qy       2401 TCCATGTCCGATTCACACGCTACACACCCACATGAACACATTCGAACTTCGAAGGCACAC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TCCATGTCCGATTCACACGCTACACACCCACATGAACACATTCGAACTTCGAAGGCACAC 2460

Qy       2461 ACTCCTGCTTCATAGGCCCCACGGTAAGTGAGTTCACACCTAGAACACTGTCCTGACCGC 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 ACTCCTGCTTCATAGGCCCCACGGTAAGTGAGTTCACACCTAGAACACTGTCCTGACCGC 2520

Qy       2521 AGGACGCGTGCCTTGGACTTGGTATTCTACATGTGACTGGCTTTCTTGCCCTCGTCTCTT 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 AGGACGCGTGCCTTGGACTTGGTATTCTACATGTGACTGGCTTTCTTGCCCTCGTCTCTT 2580

Qy       2581 GAATGTTTAGACTCTTAAGATCATATCCTG 2610
              |||||||||||||||||||||||||||||
Db       2581 GAATGTTTAGACTCTTAAGATCATATCCTG 2610
```

RESULT 2
AAH15102
ID AAH15102 standard; cDNA; 2772 BP.
XX
AC AAH15102;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:13129.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 8; SEQ ID NO 13129; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesising 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 2772 BP; 699 A; 691 C; 765 G; 617 T; 0 U; 0 Other;

Query Match 96.4%; Score 2516.2; DB 4; Length 2772;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2541; Conservative 0; Mismatches 3; Indels 3; Gaps 2;

```
Qy         65 GGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTG 124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTG 160

Qy        125 TTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCC 184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 TTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCC 220

Qy        185 AGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCAAG 244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCAAG 280

Qy        245 CTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTA 304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 CTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTA 340

Qy        305 AGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTA 364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 AGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTA 400

Qy        365 TCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCAC 424
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 TCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCAC 460

Qy        425 AGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCG 484
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 AGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCG 520

Qy        485 TGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGG 544
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        521 TGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGG 580
Qy        545 AGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCA 604
Db        581 AGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCA 640
Qy        605 GCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGG 664
Db        641 GCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGG 700
Qy        665 AACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATA 724
Db        701 AACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATA 760
Qy        725 GTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAAC 784
Db        761 GTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAAC 820
Qy        785 AGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGG 844
Db        821 AGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGG 880
Qy        845 AGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGC 904
Db        881 AGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGC 940
Qy        905 CCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACA 964
Db        941 CCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACA 1000
Qy        965 GAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACT 1024
Db       1001 GAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGAAGACT 1060
Qy       1025 TGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCC 1084
Db       1061 TGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCC 1120
Qy       1085 GCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATG 1144
Db       1121 GCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATG 1180
Qy       1145 GCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACT 1204
Db       1181 GCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACT 1240
Qy       1205 ATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAG 1264
Db       1241 ATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAG 1300
Qy       1265 ATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAG 1324
Db       1301 ATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAG 1360
Qy       1325 GAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTG 1384
Db       1361 GAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTG 1420
Qy       1385 CAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGGTG 1444
Db       1421 CAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGGTG 1480
Qy       1445 CCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTGCA 1504
Db       1481 CCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTGCA 1540
Qy       1505 GTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAAAG 1564
Db       1541 GTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAAAG 1600
Qy       1565 AACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGA 1624
Db       1601 AACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGA 1660
```

```
Qy       1625 AATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAA 1684
Db       1661 AATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAA 1720
Qy       1685 AAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGA 1744
Db       1721 AAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGA 1780
Qy       1745 AGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAG 1804
Db       1781 AGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAG 1840
Qy       1805 TAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACC 1864
Db       1841 TAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACC 1900
Qy       1865 TAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCCTGTC 1924
Db       1901 TAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCCTGTC 1960
Qy       1925 CAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCAGC 1984
Db       1961 CAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCAGC 2020
Qy       1985 GGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTCTGTC 2044
Db       2021 GGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTCTGTC 2080
Qy       2045 AGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGTGAACGTTCACCTTAAT 2104
Db       2081 AGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGTGAACGTTCACCTTAAT 2140
Qy       2105 GATTGATTCTTTAATCTCTGTTTTCACTCTCAGGCTCTGGTAAGTATTTGTATTCTCTTC 2164
Db       2141 GATTGATTCTTTAATCTCTGTTGTCACTCTCAGGCTCTGGTAAGTATTCGTATTCTCTTC 2200
Qy       2165 ATCCCAGTCTGATTGCATAGCCACACTGCCCGGCACGCCACATCCACCCCTGTCTGCACA 2224
Db       2201 ATCCCAGTCTGATTGCATAGCCACACTGCCCGGCACGCCACATCCACCCCTGTCTGCACA 2260
Qy       2225 TGAGTTGTTCTGACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGGCCAG 2284
Db       2261 TGAGTTGTTCTGACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGGCCAG 2320
Qy       2285 CACATGAAAGCATCACTTCTTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTGCATC 2344
Db       2321 CACATGAAAGCATCACTTCTTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTGCATC 2380
Qy       2345 TGATTTTCAGGTGTACATTTATTTTTGACTGGGCAGATAGGGGATTTTTTTTTTTTTCCA 2404
Db       2381 TGATTTTCAGGTGTACATTTATTTTTGACTGGGCAGATAGGGGA--TTTTTTTTTTTCCA 2438
Qy       2405 TGTCCGATTCACACGCTACACACCCACATGAACACATTCGAACTTCGAAGG-CACACACT 2463
Db       2439 TGTCCGATTCACACGCTACACACCCACATGAACACATTCGAACTTCGAAGGCCACACACT 2498
Qy       2464 CCTGCTTCATAGGCCCCACGGTAAGTGAGTTCACACCTAGAACACTGTCCTGACCGCAGG 2523
Db       2499 CCTGCTTCATAGGCCCCACGGTAAGTGAGTTCACACCTAGAACACTGTCCTGACCGCAGG 2558
Qy       2524 ACGCGTGCCTTGGACTTGGTATTCTACATGTGACTGGCTTTCTTGCCCTCGTCTCTTGAA 2583
Db       2559 ACGCGTGCCTTGGACTTGGTATTCTACATGTGACTGGCTTTCTTGCCCTCGTCTCTTGAA 2618
Qy       2584 TGTTTAGACTCTTAAGATCATATCCTG 2610
Db       2619 TGTTTAGACTCTTAAGATCATATCCTG 2645
```

RESULT 3
ABV28029
ID ABV28029 standard; cDNA; 2085 BP.

XX
AC ABV28029;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 28020.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 5788; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodyanamic or pharmacogenomic marker
XX
SQ Sequence 2085 BP; 520 A; 525 C; 657 G; 378 T; 0 U; 5 Other;

Query Match 69.4%; Score 1810.4; DB 5; Length 2085;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1960; Conservative 0; Mismatches 11; Indels 13; Gaps 13;

```
Qy          3 CGGTGGCCGTCCGCCCACTCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGCCT 62
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         17 CGGTGGCCGTCCGCCCACTCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGCCT 76

Qy         63 GCGGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCC 122
              |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         77 GCGGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAG-TGTTGAAAATACGAAGCC 135

Qy        123 TGTTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGAC 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 TGTTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGAC 195

Qy        183 CCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCA 242
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        196 CCAGAGGTTTCGTCTGCCAGGG-TTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCA 254

Qy        243 AGCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACC 302
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        255 AGCTTTCAGTG-TGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACC 313

Qy        303 TAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCAC 362
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        314 TAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCAC 373

Qy        363 TATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCC 422
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        374 TATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCG-CACCCCC 432

Qy        423 ACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCG 482
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 ACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCG 492

Qy        483 CGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCT 542
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        493 CGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCT 552

Qy        543 GGAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGG 602
              |||||||||||||||||||||||||||||||||||| |||||||||||||||||| |||||||
Db        553 GGAGGATTCTGGTGAGCCCACGTCCTGCCCCG-CACAGACGCCGAGACAG-CTCCGAGGG 610

Qy        603 CAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAA 662
              |||||||||||||||||||||||||||||||| |||||||||||||||||| ||||||
Db        611 CAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGG-CCCCAGTCTGCTTCCACAG-TGTGAA 668

Qy        663 GGAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGA 722
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        669 GGAACGACCAG-CTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGA 727

Qy        723 TAGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGA 782
              |||||||||||||||| |||||||||||||||||||||||||||||| ||||||||||||||
Db        728 TAGTGACAGCGATGG-CTGACCTTGAAAGAGCTTCAGAATCG-CTTCGCAGGAAGCGGGA 785

Qy        783 ACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCG 842
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        786 ACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCG 845

Qy        843 GGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCT 902
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 GGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCT 905

Qy        903 GCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGA 962
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 GCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGA 965

Qy        963 CAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGA 1022
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 CAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGA 1025

Qy       1023 CTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTG 1082
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 CTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTG 1085

Qy       1083 CCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCA 1142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 CCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCA 1145

Qy       1143 TGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGA 1202
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 TGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGA 1205

Qy       1203 CTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGC 1262
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1206 CTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGC 1265

Qy       1263 AGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTAT 1322
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1266 AGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTAT 1325

Qy       1323 AGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGC 1382
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1326 AGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGC 1385

Qy       1383 TGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGG 1442
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1386 TGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGG 1445

Qy       1443 TGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTG 1502
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1446 TGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTG 1505

Qy       1503 CAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAA 1562
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1506 CAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAA 1565

Qy       1563 AGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCC 1622
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1566 AGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCC 1625

Qy       1623 GAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGA 1682
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1626 GAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGA 1685

Qy       1683 AAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGG 1742
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1686 AAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGG 1745

Qy       1743 GAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGC 1802
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1746 GAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGC 1805

Qy       1803 AGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCA 1862
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1806 AGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCA 1865

Qy       1863 CCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCCTG 1922
             |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       1866 CCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAG-CATCCCCTGGGCCTG 1924

Qy       1923 TCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCA 1982
             |||||||||||||||||||| |||||||||||||||||||||||| |    | |    |
Db       1925 TCCAGGACTGGGAGTTGCAGC-TTGTGTTAAGCTGATCACAGACACGGTGCCCTCGGGGA 1983

Qy       1983 GCGG 1986
              || |
Db       1984 GCAG 1987
```

```
RESULT 4
ABV27935
ID   ABV27935 standard; cDNA; 2085 BP.
XX
AC   ABV27935;
XX
DT   16-SEP-2002  (first entry)
XX
DE   Human prostate expression marker cDNA 27926.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US005171.
XX
PR   17-FEB-2000; 2000US-0183319P.
PR   16-MAR-2000; 2000US-0189862P.
PR   25-MAY-2000; 2000US-0207454P.
PR   09-JUN-2000; 2000US-0211314P.
PR   18-JUL-2000; 2000US-0219007P.
PR   13-DEC-2000; 2000US-0255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Endege WO,  Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS   Claim 1; Page 5757-5758; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for: (a) assessing whether
CC   a patient is afflicted with prostate cancer; (b) monitoring the
CC   progression of prostate cancer in a patient; (c) assessing the efficacy
CC   of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC   the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC   determining whether prostate cancer has metastasized in a patient; (h)
CC   assessing the aggressiveness or indolence of prostate cancer in a patient
CC   ; (I) is also useful as a pharmacodyanamic or pharmacogenomic marker
XX
SQ   Sequence 2085 BP; 520 A; 525 C; 657 G; 378 T; 0 U; 5 Other;
```

```
  Query Match             69.4%;  Score 1810.4;  DB 5;  Length 2085;
  Best Local Similarity   98.8%;  Pred. No. 0;
  Matches 1960;  Conservative    0;  Mismatches   11;  Indels   13;  Gaps   13;

Qy          3 CGGTGGCCGTCCGCCCACTCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGCCT 62
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         17 CGGTGGCCGTCCGCCCACTCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGCCT 76

Qy         63 GCGGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCC 122
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         77 GCGGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAG-TGTTGAAAATACGAAGCC 135

Qy        123 TGTTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGAC 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 TGTTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGAC 195

Qy        183 CCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCA 242
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        196 CCAGAGGTTTCGTCTGCCAGGG-TTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCA 254

Qy        243 AGCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACC 302
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        255 AGCTTTCAGTG-TGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACC 313

Qy        303 TAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCAC 362
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        314 TAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCAC 373

Qy        363 TATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCC 422
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        374 TATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCG-CACCCCC 432

Qy        423 ACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCG 482
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 ACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCG 492

Qy        483 CGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCT 542
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        493 CGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCT 552

Qy        543 GGAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGG 602
              |||||||||||||||||||||||||||||||| ||||||||||||||| |||||||||
Db        553 GGAGGATTCTGGTGAGCCCACGTCCTGCCCCG-CACAGACGCCGAGACAG-CTCCGAGGG 610
```

```
Qy       603 CAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAA 662
Db       611 CAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGG-CCCCAGTCTGCTTCCACAG-TGTGAA 668

Qy       663 GGAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGA 722
Db       669 GGAACGACCAG-CTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGA 727

Qy       723 TAGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGA 782
Db       728 TAGTGACAGCGATGG-CTGACCTTGAAAGAGCTTCAGAATCG-CTTCGCAGGAAGCGGGA 785

Qy       783 ACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCG 842
Db       786 ACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCG 845

Qy       843 GGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCT 902
Db       846 GGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCT 905

Qy       903 GCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGA 962
Db       906 GCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGA 965

Qy       963 CAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGA 1022
Db       966 CAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGA 1025

Qy      1023 CTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTG 1082
Db      1026 CTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTG 1085

Qy      1083 CCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCA 1142
Db      1086 CCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCA 1145

Qy      1143 TGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGA 1202
Db      1146 TGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGA 1205

Qy      1203 CTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGC 1262
Db      1206 CTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGC 1265

Qy      1263 AGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTAT 1322
Db      1266 AGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTAT 1325

Qy      1323 AGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGC 1382
Db      1326 AGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGC 1385

Qy      1383 TGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGG 1442
Db      1386 TGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGG 1445

Qy      1443 TGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTG 1502
Db      1446 TGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTG 1505

Qy      1503 CAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAA 1562
Db      1506 CAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAA 1565

Qy      1563 AGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCC 1622
Db      1566 AGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCC 1625

Qy      1623 GAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGA 1682
Db      1626 GAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGA 1685

Qy      1683 AAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGG 1742
Db      1686 AAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGG 1745

Qy      1743 GAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGC 1802
Db      1746 GAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGC 1805

Qy      1803 AGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCA 1862
Db      1806 AGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCA 1865

Qy      1863 CCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCCTG 1922
Db      1866 CCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAG-CATCCCCTGGGCCTG 1924

Qy      1923 TCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCA 1982
Db      1925 TCCAGGACTGGGAGTTGCAGC-TTGTGTTAAGCTGATCACAGACACGGTGCCCTCGGGGA 1983

Qy      1983 GCGG 1986
Db      1984 GCAG 1987
```

RESULT 5
ABV22095
ID ABV22095 standard; cDNA; 2085 BP.
XX
AC ABV22095;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 22086.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JB;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 3785; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)

```
CC   determining whether prostate cancer has metastasized in a patient; (h)
CC   assessing the aggressiveness or indolence of prostate cancer in a patient
CC   ; (I) is also useful as a pharmacodyanamic or pharmacogenomic marker
XX
SQ   Sequence 2085 BP; 520 A; 525 C; 657 G; 378 T; 0 U; 5 Other;

  Query Match            69.4%;  Score 1810.4;  DB 5;  Length 2085;
  Best Local Similarity  98.8%;  Pred. No. 0;
  Matches 1960;  Conservative    0;  Mismatches   11;  Indels   13;  Gaps   13;

Qy          3 CGGTGGCCGTCCGCCCACTCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGCCT 62
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         17 CGGTGGCCGTCCGCCCACTCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGCCT 76

Qy         63 GCGGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCC 122
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db         77 GCGGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAG-TGTTGAAAATACGAAGCC 135

Qy        123 TGTTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGAC 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 TGTTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGAC 195

Qy        183 CCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCA 242
              |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        196 CCAGAGGTTTCGTCTGCCAGGG-TTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCA 254

Qy        243 AGCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACC 302
              ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        255 AGCTTTCAGTG-TGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACC 313

Qy        303 TAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCAC 362
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        314 TAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCAC 373

Qy        363 TATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCC 422
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        374 TATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCG-CACCCCC 432

Qy        423 ACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCG 482
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 ACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCG 492

Qy        483 CGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCT 542
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        493 CGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCT 552

Qy        543 GGAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGG 602
              |||||||||||||||||||||||||||||||| ||||||||||||||||| |||||||||
Db        553 GGAGGATTCTGGTGAGCCCACGTCCTGCCCCG-CACAGACGCCGAGACAG-CTCCGAGGG 610

Qy        603 CAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAA 662
              ||||||||||||||||||||||||||||||||| ||||||||||||||||||| ||||||
Db        611 CAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGG-CCCCAGTCTGCTTCCACAG-TGTGAA 668

Qy        663 GGAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGA 722
              ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        669 GGAACGACCAG-CTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGA 727

Qy        723 TAGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGA 782
              ||||||||||||||| |||||||||||||||||||||||||| |||||||||||||||||
Db        728 TAGTGACAGCGATGG-CTGACCTTGAAAGAGCTTCAGAATCG-CTTCGCAGGAAGCGGGA 785

Qy        783 ACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCG 842
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        786 ACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCG 845

Qy        843 GGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCT 902
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 GGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCT 905

Qy        903 GCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGA 962
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 GCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGA 965

Qy        963 CAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGA 1022
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 CAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGA 1025

Qy       1023 CTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTG 1082
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 CTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTG 1085

Qy       1083 CCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCA 1142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 CCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCA 1145

Qy       1143 TGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGA 1202
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 TGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGA 1205

Qy       1203 CTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGC 1262
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1206 CTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGC 1265

Qy       1263 AGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTAT 1322
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1266 AGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTAT 1325

Qy       1323 AGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGC 1382
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1326 AGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGC 1385

Qy       1383 TGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGG 1442
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1386 TGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGG 1445

Qy       1443 TGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTG 1502
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1446 TGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTG 1505

Qy       1503 CAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAA 1562
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1506 CAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAA 1565

Qy       1563 AGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCC 1622
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1566 AGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCC 1625

Qy       1623 GAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGA 1682
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1626 GAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGA 1685

Qy       1683 AAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGG 1742
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1686 AAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGG 1745

Qy       1743 GAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGC 1802
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1746 GAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGC 1805

Qy       1803 AGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCA 1862
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1806 AGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCA 1865

Qy       1863 CCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCCTG 1922
              |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db       1866 CCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAG-CATCCCCTGGGCCTG 1924

Qy       1923 TCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCA 1982
              ||||||||||||||||||||| |||||||||||||||||||||||| |    | |    |
Db       1925 TCCAGGACTGGGAGTTGCAGC-TTGTGTTAAGCTGATCACAGACACGGTGCCCTCGGGGA 1983

Qy       1983 GCGG 1986
              || |
Db       1984 GCAG 1987
```

```
RESULT 6
ABV22192
ID   ABV22192 standard; cDNA; 2085 BP.
XX
AC   ABV22192;
XX
DT   13-SEP-2002  (first entry)
XX
DE   Human prostate expression marker cDNA 22183.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US005171.
XX
PR   17-FEB-2000; 2000US-0183319P.
PR   16-MAR-2000; 2000US-0189862P.
PR   25-MAY-2000; 2000US-0207454P.
PR   09-JUN-2000; 2000US-0211314P.
PR   18-JUL-2000; 2000US-0219007P.
PR   13-DEC-2000; 2000US-0255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Endege WO,  Monahan JB;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS   Claim 1; Page 3816; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for: (a) assessing whether
CC   a patient is afflicted with prostate cancer; (b) monitoring the
CC   progression of prostate cancer in a patient; (c) assessing the efficacy
CC   of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC   the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC   determining whether prostate cancer has metastasized in a patient; (h)
CC   assessing the aggressiveness or indolence of prostate cancer in a patient
CC   ; (I) is also useful as a pharmacodyanamic or pharmacogenomic marker
XX
SQ   Sequence 2085 BP; 520 A; 526 C; 656 G; 378 T; 0 U; 5 Other;

  Query Match             69.3%;  Score 1808.8;  DB 5;  Length 2085;
  Best Local Similarity   98.7%;  Pred. No. 0;
  Matches 1959;  Conservative    0;  Mismatches   12;  Indels   13;  Gaps   13;

Qy          3 CGGTGGCCGTCCGCCCACTCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGCCT 62
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         17 CGGTGGCCGTCCGCCCACTCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGCCT 76

Qy         63 GCGGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCC 122
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db         77 GCGGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAG-TGTTGAAAATACGAAGCC 135

Qy        123 TGTTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGAC 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 TGTTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGAC 195

Qy        183 CCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCA 242
              |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        196 CCAGAGGTTTCGTCTGCCAGGG-TTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCA 254

Qy        243 AGCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACC 302
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        255 AGCTTTCAGTG-TGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACC 313

Qy        303 TAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCAC 362
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        314 TAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCAC 373

Qy        363 TATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCC 422
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        374 TATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCG-CACCCCC 432

Qy        423 ACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCG 482
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 ACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCG 492

Qy        483 CGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCT 542
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        493 CGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCT 552

Qy        543 GGAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGG 602
              |||||||||||||||||||||||||||||| |||||||||||||||| |||||||||
Db        553 GGAGGATTCTGGTGAGCCCACGTCCTGCCCCG-CACAGACGCCGAGACAG-CTCCGAGGG 610

Qy        603 CAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAA 662
              |||||||||||||||||||||||||||||| |||||||||||||||| ||||||
Db        611 CAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGG-CCCCAGTCTGCTTCCACAG-TGTGAA 668

Qy        663 GGAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGA 722
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        669 GGAACGACCAG-CTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGA 727

Qy        723 TAGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGA 782
              |||||||||||||| |||||||||||||||||||||||| ||||||||||||||||||
Db        728 TAGTGACAGCGATGG-CTGACCTTGAAAGAGCTTCAGAATCG-CTTCGCAGGAAGCGGGA 785

Qy        783 ACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCG 842
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        786 ACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCG 845

Qy        843 GGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCT 902
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 GGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCT 905

Qy        903 GCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGA 962
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 GCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGA 965

Qy        963 CAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGA 1022
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 CAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGA 1025

Qy       1023 CTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTG 1082
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 CTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTG 1085

Qy       1083 CCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCA 1142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 CCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCA 1145

Qy       1143 TGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGA 1202
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 TGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGA 1205

Qy       1203 CTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGC 1262
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1206 CTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGC 1265

Qy       1263 AGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTAT 1322
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1266 AGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTAT 1325

Qy       1323 AGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGC 1382
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1326 AGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGC 1385

Qy       1383 TGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGG 1442
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1386 TGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGG 1445

Qy       1443 TGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTG 1502
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1446 TGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTG 1505

Qy       1503 CAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAA 1562
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1506 CAGTAATGACTGTATCCTCAAACACGCCGCACCGACAATGAAGTTTCTAAGCTCAGGTAA 1565

Qy       1563 AGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCC 1622
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1566 AGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCC 1625

Qy       1623 GAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGA 1682
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1626 GAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGA 1685

Qy       1683 AAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGG 1742
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1686 AAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGG 1745

Qy       1743 GAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGC 1802
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1746 GAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGC 1805

Qy       1803 AGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCA 1862
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1806 AGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCA 1865

Qy       1863 CCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCCTG 1922
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       1866 CCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAG-CATCCCCTGGGCCTG 1924

Qy       1923 TCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCA 1982
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db       1925 TCCAGGACTGGGAGTTGCAGC-TTGTGTTAAGCTGATCACAGACACGGTGCCCTCGGGGA 1983

Qy       1983 GCGG 1986
              |||| |
Db       1984 GCAG 1987
```

```
RESULT 7
AAC77933
ID   AAC77933 standard; cDNA; 1764 BP.
XX
AC   AAC77933;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human cancer associated gene sequence SEQ ID NO:327.
XX
KW   Human; cancer associated gene; cancer antigen; detection; cancer;
KW   diagnosis; cytostatic; proliferative; vulnerary; immunomodulator;
KW   antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW   antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW   dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW   vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW   immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW   allergic reaction; graft versus host disease; organ rejection;
KW   haemostatic; thrombolytic; cardiovascular disorder; infection;
KW   neurological disease; drug screening; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055350-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US005882.
XX
PR   12-MAR-1999;   99US-0124270P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587533/55.
DR   P-PSDB; AAB43724.
XX
PT   Novel isolated nucleic acids comprising sequences encoding peptides
PT   useful for treating or diagnosing e.g. cancer.
XX
PS   Claim 1; Page 880-881; 2352pp; English.
XX
CC   AAC77607 to AAC78448 encode the human cancer associated proteins given in
CC   AAB43398 to AAB44239. The proteins can have activities based on the
CC   tissues and cells the genes are expressed in. Example of activities
CC   include: cytostatic; proliferative; vulnerary; immunomodulator;
CC   antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC   antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC   dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC   nootropic; vasotropic; antipsoriatic and antiangiogenic. The
CC   polynucleotides and polypeptides can be used for preventing, treating or
CC   ameliorating medical conditions and diagnosing pathological conditions.
CC   Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC   the present invention may be used to treat immune disorders by activating
CC   or inhibiting the proliferation, differentiation or mobilisation of
CC   immune cells, to treat disorders of haematopoietic cells, autoimmune
CC   disorders, allergic reactions, graft versus host disease and organ
CC   rejection, modulate haemostatic or thrombolytic activity, modulate
CC   inflammation, cancers, cardiovascular disorders, neurological disease and
CC   bacterial or viral infections. The peptides, nucleotides, antibodies,
CC   agonists and antagonists may be also be used in drug screens. AAC78449 to
CC   AAC78457 and AAB44240 represent sequences used in the exemplification of
CC   the present invention
XX
SQ   Sequence 1764 BP; 468 A; 424 C; 415 G; 453 T; 0 U; 4 Other;
```

Query Match 59.9%; Score 1564.4; DB 3; Length 1764;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1610; Conservative 0; Mismatches 2; Indels 5; Gaps 4;

```
Qy        996 GGACATCAAAGATGAGGAGCCTGGAGAC-TTGGGCCGACCGAAGCCTGAATGTGAGGGTT 1054
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db          1 GGACATCAAAGATGAGGAGCCTGGAGACTTTGGGCCGACCGAAGCCTGAATGTGAGGGTT 60

Qy       1055 ACGACCCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTT 1114
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACGACCCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTT 120

Qy       1115 GCTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAG 1174
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAG 180

Qy       1175 GGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACTGCACCATTCTGCAAG 1234
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACTGCACCATTCTGCAAG 240

Qy       1235 TGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAG 1294
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAG 300

Qy       1295 ATGCTGATGGCACCGATTGTACAAGTATAGGAACAATAGAGCAGAAGTCTAGCGAAGACC 1354
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ATGCTGATGGCACCGATTGTACAAGTATAGGAACAATAGAGCAGAAGTCTAGCGAAGACC 360
```

```
Qy       1355 AAGGGATAAAGGGTAGAATTGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGA 1414
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAGGGATAAAGGGTAGAATTGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGA 420

Qy       1415 TCTTCCAGCCTGTGATAGAGGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTC 1474
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCTTCCAGCCTGTGATAGAGGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTC 480

Qy       1475 ACGTGGCGCAGCCCGACTCGGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAG 1534
             |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ACGTGGCGCA-CCCGACTCGGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAG 539

Qy       1535 CGACAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGA 1594
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 CGACAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGA 599

Qy       1595 AGATGAAGCCAGAGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCT 1654
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 AGATGAAGCCAGAGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCT 659

Qy       1655 CTTCTGTGCACAAGAGACCAGCTCCAGAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGG 1714
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 CTTCTGTGCACAAGAGACCAGCTCCAGAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGG 719

Qy       1715 TGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGT 1774
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 TGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGT 779

Qy       1775 CGTGGGCGAGCGATCACAATTACAATGCAGTAAAGCCAGAAAAGACTGCTGCTCCCTCGC 1834
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        780 CGTGGGCGAGCGATCACAATTACAATGCAGTAAAGCCAGAAAAGACTGCTGCTCCCTCGC 839

Qy       1835 CGTCACTGTTGTATAAATGTATGTATCACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTT 1894
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 CGTCACTGTTGTATAAATGTATGTATCACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTT 899

Qy       1895 CTTTCTGGATAGCCATCCCCTGGGCCTGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAG 1954
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        900 CTTTCTGGATAGCCATCCCCTGGGCCTGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAG 959

Qy       1955 CTGATCACAGACACCGGCTGCACCATCAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTC 2014
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 CTGATCACAGACACCGGCTGCACCATCAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTC 1019

Qy       2015 CTGCTGCCCTGTGTCCATCCCTAGTCTGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTG 2074
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1020 CTGCTGCCCTGTGTCCATCCCTAGTCTGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTG 1079

Qy       2075 TAAACCTCACTGGTGAACGTTCACCTTAATGATTGATTCTTTAATCTCTGTTTTCACTCT 2134
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 TAAACCTCACTGGTGAACGTTCACCTTAATGATTGATTCTTTAATCTCTGTTTTCACTCT 1139

Qy       2135 CAGGCTCTGGTAAGTATTTGTATTCTCTTCATCCCAGTCTGATTGCATAGCCACACTGCC 2194
             |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       1140 CAGGCTCTGGTAAGTATTCGTATTCTCTTCATCCCAGTCTGATTGCATAGCCACACTGCC 1199

Qy       2195 CGGCACGCCACATCCACCCCTGTCTGCACATGAGTTGTTCTGACAACAGCGCTGTATACG 2254
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1200 CGGCACGCCACATCCACCCCTGTCTGCACATGAGTTGTTCTGACAACAGCGCTGTATACG 1259

Qy       2255 CTTCAGTTTTTCCACATTGTCCACGGCCAGCACATGAAAGCATCACTTCTTTTTTATGTT 2314
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1260 CTTCAGTTTTTCCACATTGTCCACGGCCAGCACATGAAAGCATCACTTCTTTTTTATGTT 1319

Qy       2315 GTGGGAATCTTTGCAAGTTAGTGTTGCATCTGATTTTCAGGTGTACATTTATTTTTGACT 2374
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1320 GTGGGAATCTTTGCAAGTTAGTGTTGCATCTGATTTTCAGGTGTACATTTATTTTTGACT 1379

Qy       2375 GGGCAGATAGGGGATTTTTTTTTTTTTCCATGTCCGATTCACACGCTACACACCCACATG 2434
             ||||||||||||||  |||| ||||||||||||||||||||||||||||||||||||||||||
Db       1380 GGGCAGATAGGGGA--TTTTNTTTTTTCCATGTCCGATTCACACGCTACACACCCACATG 1437

Qy       2435 AACACATTCGAACTTCGAAGG-CACACACTCCTGCTTCATAGGCCCCACGGTAAGTGAGT 2493
             ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       1438 AACACATTCGAACTTCGAAGGCCACACACTCCTGCTTCATAGGCCCCACGGTAAGTGAGT 1497

Qy       2494 TCACACCTAGAACACTGTCCTGACCGCAGGACGCGTGCCTTGGACTTGGTATTCTACATG 2553
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1498 TCACACCTAGAACACTGTCCTGACCGCAGGACGCGTGCCTTGGACTTGGTATTCTACATG 1557

Qy       2554 TGACTGGCTTTCTTGCCCTCGTCTCTTGAATGTTTAGACTCTTAAGATCATATCCTG 2610
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1558 TGACTGGCTTTCTTGCCCTCGTCTCTTGAATGTTTAGACTCTTAAGATCATATCCTG 1614
```

RESULT 8
AAK53352
ID AAK53352 standard; cDNA; 7290 BP.
XX
AC AAK53352;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 2881.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US004098.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR P-PSDB; AAM80219.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.
XX
PS Claim 1; Page 5053-5055; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
SQ Sequence 7290 BP; 1762 A; 1953 C; 1975 G; 1600 T; 0 U; 0 Other;

```
  Query Match             49.1%;  Score 1280.6;  DB 4;  Length 7290;
  Best Local Similarity   92.6%;  Pred. No. 0;
  Matches 1401;  Conservative    0;  Mismatches    4;  Indels  108;  Gaps    1;

Qy       455 GTGGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCC 514
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        64 GTGGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCC 123

Qy       515 GCGGCAGGAGGAGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCG 574
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       124 GCGGCAGGAGGAGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCG 183

Qy       575 CCACAGACGCCGAGACAGCCTCCGAGGGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCG 634
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       184 CCACAGACGCCGAGACAGCCTCCGAGGGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCG 243

Qy       635 GCCCCCAGTCTGCTTCCACAGCTGTGAAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAG 694
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       244 GCCCCCAGTCTGCTTCCACAGCTGTGAAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAG 303

Qy       695 GAGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGGCCTGACCTTGAAAGAGC 754
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       304 GAGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGGCCTGACCTTGAAAGAGC 363

Qy       755 TTCAGAATCGCCTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGA 814
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       364 TTCAGAATCGCCTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGA 423

Qy       815 TCCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCG 874
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       424 TCCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCG 483

Qy       875 AGGCCAGTGACACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGG 934
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       484 AGGCCAGTGACACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGG 543

Qy       935 GGGTTGTGTCCCAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTC 994
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       544 GGGTTGTGTCCCAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTC 603

Qy       995 AGGACATCAAAGATGAGGAGCCTGGAGACTTGGGCCGACCGAAGCCTGAATGTGAGGGTT 1054
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       604 AGGACATCAAAGATGAGGAGCCTGGAGACTTGGGCCGACCGAAGCCTGAATGTGAGGGTT 663

Qy      1055 ACGACCCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTT 1114
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       664 ACGACCCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTT 723

Qy      1115 GCTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAG 1174
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       724 GCTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAG 783

Qy      1175 GGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACTGCACCATTCTGCAAG 1234
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       784 GGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACTGCACCATTCTGCAAG 843

Qy      1235 TGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAG 1294
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       844 TGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAG 903

Qy      1295 ATGCTGATGGCACCGATTGTACAAGTATAGGAACAATAGAGCAGAAGTCTAGCGAAGACC 1354
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       904 ATGCTGATGGCACCGATTGTACAAGTATAGGAACAATAGAGCAGAAGTCTAGCGAAGACC 963

Qy      1355 AAGGGATAAAGGGTAGAATTGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGA 1414
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       964 AAGGGATAAAGGGTAGAATTGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGA 1023

Qy      1415 TCTTCCAGCCT---------------------------------------------------- 1425
            ||||||||||||
Db      1024 TCTTCCAGCCTGGTCCCGGGCCTGTCCCCACCCAGCTGCCTGTCCTCTGGCAGGTATTGG 1083

Qy      1426 ----------------------------------------------------------G 1426
                                                                        |
Db      1084 AAATTGCTGTGTCTAGAAGCATCTCTGCCTTCACACTCTTGCACTGCATTAGCTGCAAGG 1143

Qy      1427 TGATAGAGGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGC 1486
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1144 TGATAGAGGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGC 1203

Qy      1487 CCGACTCGGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGT 1546
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1204 CCGACTCGGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGT 1263

Qy      1547 TTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAG 1606
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1264 TTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAG 1323

Qy      1607 AGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACA 1666
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1324 AGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACA 1383

Qy      1667 AGAGACCAGCTCCAGAAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGC 1726
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1384 AGAGACCAGCTCCAGAAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGC 1443

Qy      1727 GGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCG 1786
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1444 GGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCG 1503

Qy      1787 ATCACAATTACAATGCAGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGT 1846
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1504 ATCACAATTACAATGCAGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGT 1563

Qy      1847 ATAAATGTATGTA 1859
            ||||||  | | |
Db      1564 ATAAATCCACGAA 1576
```

```
RESULT 9
AAK52368
ID   AAK52368 standard; cDNA; 7838 BP.
XX
AC   AAK52368;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 913.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US004098.
XX
PR   03-FEB-2000; 2000US-00496914.
PR   27-APR-2000; 2000US-00560875.
PR   20-JUN-2000; 2000US-00598075.
PR   19-JUL-2000; 2000US-00620325.
PR   01-SEP-2000; 2000US-00654936.
PR   15-SEP-2000; 2000US-00663561.
PR   20-OCT-2000; 2000US-00693325.
PR   30-NOV-2000; 2000US-00728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y;
PI   Ma Y,  Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
```

```
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   P-PSDB; AAM79235.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT   in diagnosis and gene therapy.
XX
PS   Claim 1; Page 3007-3013; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC   (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC   sequence listing were missing at the time of publication
XX
SQ   Sequence 7838 BP; 1872 A; 2123 C; 2118 G; 1725 T; 0 U; 0 Other;

  Query Match             49.1%;  Score 1280.6;  DB 4;  Length 7838;
  Best Local Similarity   92.6%;  Pred. No. 0;
  Matches 1401;  Conservative    0;  Mismatches    4;  Indels  108;  Gaps    1;

Qy        455 GTGGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCC 514
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 GTGGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCC 658

Qy        515 GCGGCAGGAGGAGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCG 574
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 GCGGCAGGAGGAGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCG 718

Qy        575 CCACAGACGCCGAGACAGCCTCCGAGGGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCG 634
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 CCACAGACGCCGAGACAGCCTCCGAGGGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCG 778

Qy        635 GCCCCCAGTCTGCTTCCACAGCTGTGAAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAG 694
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        779 GCCCCCAGTCTGCTTCCACAGCTGTGAAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAG 838

Qy        695 GAGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGGCCTGACCTTGAAAGAGC 754
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 GAGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGGCCTGACCTTGAAAGAGC 898

Qy        755 TTCAGAATCGCCTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGA 814
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        899 TTCAGAATCGCCTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGA 958

Qy        815 TCCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCG 874
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        959 TCCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCG 1018

Qy        875 AGGCCAGTGACACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGG 934
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 AGGCCAGTGACACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGG 1078

Qy        935 GGGTTGTGTCCCAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTC 994
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1079 GGGTTGTGTCCCAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTC 1138

Qy        995 AGGACATCAAAGATGAGGAGCCTGGAGACTTGGGCCGACCGAAGCCTGAATGTGAGGGTT 1054
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 AGGACATCAAAGATGAGGAGCCTGGAGACTTGGGCCGACCGAAGCCTGAATGTGAGGGTT 1198

Qy       1055 ACGACCCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTT 1114
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1199 ACGACCCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTT 1258

Qy       1115 GCTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAG 1174
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1259 GCTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAG 1318

Qy       1175 GGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACTGCACCATTCTGCAAG 1234
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1319 GGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACTGCACCATTCTGCAAG 1378

Qy       1235 TGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAG 1294
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1379 TGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAG 1438

Qy       1295 ATGCTGATGGCACCGATTGTACAAGTATAGGAACAATAGAGCAGAAGTCTAGCGAAGACC 1354
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1439 ATGCTGATGGCACCGATTGTACAAGTATAGGAACAATAGAGCAGAAGTCTAGCGAAGACC 1498

Qy       1355 AAGGGATAAAGGGTAGAATTGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGA 1414
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1499 AAGGGATAAAGGGTAGAATTGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGA 1558

Qy       1415 TCTTCCAGCCT---------------------------------------------------- 1425
              |||||||||||
Db       1559 TCTTCCAGCCTGGTCCCGGGCCTGTCCCCACCCAGCTGCCTGTCCTCTGGCAGGTATTGG 1618

Qy       1426 ----------------------------------------------------------G 1426
                                                                         |
Db       1619 AAATTGCTGTGTCTAGAAGCATCTCTGCCTTCACACTCTTGCACTGCATTAGCTGCAAGG 1678

Qy       1427 TGATAGAGGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGC 1486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1679 TGATAGAGGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGC 1738

Qy       1487 CCGACTCGGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGT 1546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1739 CCGACTCGGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGT 1798

Qy       1547 TTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAG 1606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1799 TTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAG 1858

Qy       1607 AGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACA 1666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1859 AGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACA 1918

Qy       1667 AGAGACCAGCTCCAGAAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGC 1726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1919 AGAGACCAGCTCCAGAAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGC 1978

Qy       1727 GGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCG 1786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1979 GGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCG 2038

Qy       1787 ATCACAATTACAATGCAGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGT 1846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2039 ATCACAATTACAATGCAGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGT 2098

Qy       1847 ATAAATGTATGTA 1859
              |||||| | | |
Db       2099 ATAAATCCACGAA 2111


RESULT 10
AAS86576
ID   AAS86576 standard; cDNA; 2332 BP.
XX
AC   AAS86576;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #22380.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
```

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG22389.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 22380; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2332 BP; 511 A; 621 C; 764 G; 436 T; 0 U; 0 Other;

Query Match 45.6%; Score 1190.8; DB 5; Length 2332;
Best Local Similarity 92.2%; Pred. No. 4.3e-310;
Matches 1362; Conservative 0; Mismatches 37; Indels 79; Gaps 7;

```
Qy        244 GCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCT 303
              |||| || | |      ||||||||||||||||||||||||||||||||||||||||||||||
Db        370 GCTTACACTTTACCGCTAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCT 429

Qy        304 AAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        430 AAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACT 489

Qy        364 ATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCA 423
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        490 ATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCA 549

Qy        424 CAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGC 483
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        550 CAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGC 609

Qy        484 GTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTG 543
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        610 GTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTG 669

Qy        544 GAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGC 603
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        670 GAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGC 729

Qy        604 AGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAG 663
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        730 AGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAG 789

Qy        664 GAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGAT 723
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        790 GAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGAT 849

Qy        724 AGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAA 783
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        850 AGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAA 909

Qy        784 CAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGG 843
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        910 CAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGG 969

Qy        844 GAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTG 903
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        970 GAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTG 1029

Qy        904 CCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGAC 963
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1030 CCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGAC 1089

Qy        964 AGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGAC 1023
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1090 AGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGAC 1149

Qy        1024 -TTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTG 1082
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1150 TTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTG 1209

Qy        1083 CCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCA 1142
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1210 CCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCA 1269

Qy        1143 TGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGA 1202
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1270 TGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGA 1329

Qy        1203 CTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGC 1262
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1330 CTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGC 1389

Qy        1263 AGATCAGCAGGAAGCTAAATGGAGA---CCTGGAGATGC-TGATGGCACCGA-TTGTACA 1317
               |||||||||||||||| ||  | ||   | |||||||| |||||||||||| || ||||
Db        1390 AGATCAGCAGGAAGCTTAAATGGGAGACCTGGGAGATGCTTGATGGCACCGATTTTTACA 1449

Qy        1318 AGTATAGGAACAAT--AGAGCAGAAGTCT----AGCGAAGACCAAGGGATAAAGGGTAGA 1371
               |||||||| |  ||  ||||||||||||     |   |  | | ||| ||||||||||||
Db        1450 AGTATAGGGAACATTAGGAGCAGAAGTCTTAGCCGAAGACCCAAGGGGTTAAAGGGTAGA 1509

Qy        1372 ATTGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCC------- 1424
               |||||||||||||||||||||||||||||||| |||||||||||| |   |
Db        1510 ATTGAGAAAGCTGCAAATCCAAGTGGCAAGAAGGAACTCAAGATCTTTCCAGCCTGTATT 1569

Qy        1425 ---------------------------------------------------------- 1424

Db        1570 GGAAATTGCTGTGTCTAGAAGCATCTCTGCCTTCACACTCTTGCACTGCATTAGCTGCAA 1629

Qy        1425 TGTGATAGAGGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCA 1484
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1630 GGTGATAGAGGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCA 1689

Qy        1485 GCCCGACTCGGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAA 1544
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1690 GCCCGACTCGGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAA 1749
```

```
Qy       1545 GTTTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCC 1604

Db       1750 GTTTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCC 1809

Qy       1605 AGAGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGGCAG 1642

Db       1810 AGAGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGCCTG 1847
```

```
RESULT 11
AAZ90579
ID   AAZ90579 standard; cDNA; 2867 BP.
XX
AC   AAZ90579;
XX
DT   19-JUN-2000  (first entry)
XX
DE   Murine death inducer-obliterator 1 (DIO-1) polypeptide encoding cDNA.
XX
KW   Death inducer-obliterator 1; DIO-1; cell death; cancer; tumour; murine;
KW   autoimmune disease; cytostatic; immunosuppressive; antidiabetic;
KW   antirheumatic; antiinflammatory; antiproliferative; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             232. .2076
FT                   /*tag=  a
FT                   /product= "DIO-1"
XX
PN   WO200015787-A1.
XX
PD   23-MAR-2000.
XX
PF   10-SEP-1999;   99WO-GB003019.
XX
PR   10-SEP-1998;   98SE-00003069.
PR   17-SEP-1998;   98US-0100873P.
XX
PA   (CNSJ ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.
PA   (BANN/) BANNERMAN D G.
XX
PI   Alonso CM,  Domingo DG,  Grandien A,  Leonardo E,  Martinez P;
XX
DR   WPI; 2000-271426/23.
DR   P-PSDB; AAY67580.
XX
PT   New DNA encoding human and murine death inducer-obliterator 1
PT   polypeptides, useful in the treatment of cancer, autoimmune diseases,
PT   diabetes, rheumatoid arthritis, benign tumors, malignant tumors and
PT   hyperproliferative skin disorders.
XX
PS   Claim 3; Fig 1B; 27pp; English.
XX
CC   The invention provides nucleic acids encoding the human and murine death
CC   inducer-obliterator 1 (DIO-1) polypeptides. The polypeptides can be
CC   expressed by standard recombinant methodology. The DIO-1 polypeptides,
CC   agonists and antagonists are used as a medicament for treating diseases
CC   characterized by an alteration in cell death or by hyperproliferation,
CC   e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign
CC   tumours, malignant tumours or hyperproliferative skin disorders. They are
CC   also useful in the treatment of metabolic, proliferative or inflammatory
CC   conditions. The present sequence represents a cDNA encoding the murine
CC   DIO-1 polypeptide
XX
SQ   Sequence 2867 BP; 763 A; 666 C; 774 G; 664 T; 0 U; 0 Other;

  Query Match             39.2%;  Score 1023.6;  DB 3;  Length 2867;
  Best Local Similarity   71.9%;  Pred. No. 5.9e-265;
  Matches 1494;  Conservative    0;  Mismatches  534;  Indels   50;  Gaps   10;
```

```
Qy         21 TCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGCCTGCGGAGCTTACTCCACGG 80

Db          1 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 60

Qy         81 GAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTGTTACTCGTGAACAGTG 140

Db         61 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 120

Qy        141 GCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCCAGAGGTTTCGTCTGCC 200

Db        121 ACTGACAACAGTGGGGTGAG--GCTTGGCCGTCTGCTTGACC-----------TGGCCC 166

Qy        201 AGGGTTTTTGGTTGTATTTAGGATTT-CAGGGAAAAGTGTCCAAGCTTTCAGTGTTGGAG 259

Db        167 CAGGTCTATAATTTTATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGA 226

Qy        260 CAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTAAGGCCATCAAACCCA 319

Db        227 CAGGTATGGATGATAAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCA 286

Qy        320 CCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTATCGCCAAGCGAGAGG 379

Db        287 CCAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGG 346

Qy        380 GCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCACAGCAGCAGCTGGGCC 439

Db        347 GTGCAGGAGACACGGAGGCGGACCCCAGTGAGCAGCAACC------ACAGCAGCATAACC 400

Qy        440 TGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGA 499

Db        401 TCTCCCTGCGCCGCAGTGGACGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTA 460

Qy        500 CCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGC 559

Db        461 CCACGGTTCGGCGCCGAGGGAAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGC 520

Qy        560 CCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCAGCGTGGAAAGCGCTT 619

Db        521 CCACATCTTCCACAGTCACTGATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTT 580

Qy        620 CTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGGAACGACCAGCCTCTT 679

Db        581 CTGAGATCAGAAGTGGCCCTGTATCTGACTCCTTA---GGGAAAGAACATCCTGCCTCTT 637

Qy        680 CTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGGCC 739

Db        638 CTGAAAAGGCAAAAGGAGGTGAAGAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCC 697

Qy        740 TGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGA 799

Db        698 TTACGTTGAAGGAACTTCAGAACCGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGA 757

Qy        800 GGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCG 859

Db        758 GGTCCCTGAGAGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCG 817

Qy        860 AGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGC 919

Db        818 AAACTGGGAGTGTCCAAATAGGCAGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGC 877

Qy        920 CCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGG 979

Db        878 CTGAGGCTAGTCAGGGACCAGTGTCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGG 937

Qy        980 AGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACTTGGGCCGACCGAAGC 1039

Db        938 AAGGGAAGGCGACTCAGGGAAATACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGC 997

Qy       1040 CTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACA 1099

Db        998 CTGAGTGTGAGGTTTACGACCCCAATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACA 1057

Qy       1100 ACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCA 1159
```

```
Db       1058 ACAGGTTTATGATCTGCTGTGATCGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTA 1117

Qy       1160 TTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACT 1219

Db       1118 TTTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATT 1177

Qy       1220 GCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTA 1279

Db       1178 GCACCATTTTGCAAGTGCAGGATGAGACAAACGGTAGCGCCACCAATGAGCAGGACTCTG 1237

Qy       1280 AATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAGGAACAATAGAGCAGA 1339

Db       1238 GGTGCAGATCTGTGGGTGCTGATGGCACAGACTGCACAAGCATAGGGACAGTAGAGCAGA 1297

Qy       1340 AGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTGCAAATCCAAGTGGCA 1399

Db       1298 AGTCCGGAGAAGACCAGGGCATAAAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCA 1357

Qy       1400 AGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGGTGCCTCAAAATGTATTG 1459

Db       1358 AGAAAAAACTCAAGATATTCCAGCCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTG 1417

Qy       1460 GCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTGCAGTAATGACTGTATCC 1519

Db       1418 GCCCTGGGTGTTCCAGTGTAGCACAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTC 1477

Qy       1520 TCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGC 1579

Db       1478 TCAAACACGCAGCAGCTACCATGAGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAAC 1537

Qy       1580 CTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGG 1639

Db       1538 CCAAGGAAAAGGTCAAGACGAAGCCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGG 1597

Qy       1640 CAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAAAAAAAGAGACCACAG 1699

Db       1598 TGGGGATTAAAATCTCTTCTGTGCACAAGAGACTAGCGTCAGAGAAAAGGGAAAACCCAG 1657

Qy       1700 TGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTG 1759

Db       1658 TGAAGAA---AGTGATGCTGGCTTCCAGGAGTGAGACTTCTGGGAAGGAGGCAGCCTGTG 1714

Qy       1760 AGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAGTAAAGCCAGAAAAGA 1819

Db       1715 AGAGCAGCACACCATCCTGGGCAAGTGACCACAACTACAATGCTGTGAAGCCAGAGAAGC 1774

Qy       1820 CTG---------CTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACCTAGGGG 1870

Db       1775 CAGAGAAGCCCACTGCACTCTCGCCCACCCTATTGAGTAAATGTACGTATCACCCAAAGG 1834

Qy       1871 TTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCCTGTCCAGGAC 1930

Db       1835 CTGGCTTCCCAGGCCCCTCCCATCATCT-----GGGTGGCTGCCTGGGGCTGTCTAGGAC 1889

Qy       1931 TGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCAGCGGGAAG 1990

Db       1890 CAG-----AGTCCTGGGTGTTCTGGTGCTGATAGTAGCCAGCAGC----TCACTGCCAGC 1940

Qy       1991 CAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTCTGTCAGGACT 2050

Db       1941 CAGAAGCAGATACCAAGATGCCTCTGGACCCCAGGTGTTCCTGCCTAGCCTGTGGAGCCT 2000

Qy       2051 TCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGT 2088

Db       2001 CTCTGGGTGGTTCCTAAAGAGCTGTGTAGGCCTCATGT 2038


RESULT 12
AAK67051
ID   AAK67051 standard; DNA; 30626 BP.
XX
AC   AAK67051;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:21863.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
```

PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 21863; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 30626 BP; 7102 A; 7610 C; 7817 G; 8097 T; 0 U; 0 Other;

Query Match 36.3%; Score 948.4; DB 4; Length 30626;
Best Local Similarity 99.6%; Pred. No. 3.7e-244;
Matches 972; Conservative 0; Mismatches 1; Indels 3; Gaps 2;

```
Qy      1636 CAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAAAAAAAAGAGACC 1695
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4785 CAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAAAAAAAAGAGACC 4844

Qy      1696 ACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCT 1755
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4845 ACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCT 4904

Qy      1756 TGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAGTAAAGCCAGAA 1815
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4905 TGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAGTAAAGCCAGAA 4964

Qy      1816 AAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACCTAGGGGTTGGC 1875
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4965 AAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACCTAGGGGTTGGC 5024

Qy      1876 CTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCCTGTCCAGGACTGGGA 1935
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5025 CTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCCTGTCCAGGACTGGGA 5084

Qy      1936 GTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCAGCGGGAAGCAGAG 1995
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5085 GTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCAGCGGGAAGCAGAG 5144

Qy      1996 CCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTCTGTCAGGACTTCCTG 2055
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5145 CCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTCTGTCAGGACTTCCTG 5204

Qy      2056 TCACTGTTTTCCAAAGCTGTAAACCTCACTGGTGAACGTTCACCTTAATGATTGATTCTT 2115
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5205 TCACTGTTTTCCAAAGCTGTAAACCTCACTGGTGAACGTTCACCTTAATGATTGATTCTT 5264
```

```
Qy       2116 TAATCTCTGTTTTCACTCTCAGGCTCTGGTAAGTATTTGTATTCTCTTCATCCCAGTCTG 2175
             ||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db       5265 TAATCTCTGTTTTCACTCTCAGGCTCTGGTAAGTATTCGTATTCTCTTCATCCCAGTCTG 5324

Qy       2176 ATTGCATAGCCACACTGCCCGGCACGCCACATCCACCCCTGTCTGCACATGAGTTGTTCT 2235
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5325 ATTGCATAGCCACACTGCCCGGCACGCCACATCCACCCCTGTCTGCACATGAGTTGTTCT 5384

Qy       2236 GACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGGCCAGCACATGAAAGC 2295
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5385 GACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGGCCAGCACATGAAAGC 5444

Qy       2296 ATCACTTCTTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTGCATCTGATTTTCAGG 2355
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5445 ATCACTTCTTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTGCATCTGATTTTCAGG 5504

Qy       2356 TGTACATTTATTTTTGACTGGGCAGATAGGGGATTTTTTTTTTTTTCCATGTCCGATTCA 2415
             |||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db       5505 TGTACATTTATTTTTGACTGGGCAGATAGGGGA--TTTTTTTTTTTCCATGTCCGATTCA 5562

Qy       2416 CACGCTACACACCCACATGAACACATTCGAACTTCGAAGG-CACACACTCCTGCTTCATA 2474
             |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       5563 CACGCTACACACCCACATGAACACATTCGAACTTCGAAGGCCACACACTCCTGCTTCATA 5622

Qy       2475 GGCCCCACGGTAAGTGAGTTCACACCTAGAACACTGTCCTGACCGCAGGACGCGTGCCTT 2534
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5623 GGCCCCACGGTAAGTGAGTTCACACCTAGAACACTGTCCTGACCGCAGGACGCGTGCCTT 5682

Qy       2535 GGACTTGGTATTCTACATGTGACTGGCTTTCTTGCCCTCGTCTCTTGAATGTTTAGACTC 2594
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5683 GGACTTGGTATTCTACATGTGACTGGCTTTCTTGCCCTCGTCTCTTGAATGTTTAGACTC 5742

Qy       2595 TTAAGATCATATCCTG 2610
             ||||||||||||||||
Db       5743 TTAAGATCATATCCTG 5758
```

```
RESULT 13
AAS86575
ID   AAS86575 standard; cDNA; 2986 BP.
XX
AC   AAS86575;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #22379.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG22388.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 22379; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequence of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 2986 BP; 798 A; 796 C; 781 G; 611 T; 0 U; 0 Other;
```

```
  Query Match             26.2%;  Score 682.6;  DB 5;  Length 2986;
  Best Local Similarity   88.1%;  Pred. No. 5.3e-173;
  Matches  891;  Conservative    0;  Mismatches    4;  Indels  116;  Gaps    9;

Qy        965 GAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACT 1024
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACT 60

Qy       1025 TGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCC 1084
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCC 120

Qy       1085 GCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATG 1144
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATG 180

Qy       1145 GCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACT 1204
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACT 240

Qy       1205 ATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAG 1264
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAG 300

Qy       1265 ATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAG 1324
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAG 360

Qy       1325 GAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTG 1384
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTG 420

Qy       1385 CAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCT---------------------- 1425
              ||||||||||||||||||||||||||||||||||||||||
Db        421 CAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGGTCCCGGGCCTGTCCCCA 480

Qy       1426 ---------------------------------------------------------- 1425

Db        481 CCCAGCTGCCTGTCCTCTGGCAGGTATTGGAAATTGCTGTGTCTAGAAGCATCTCTGCCT 540

Qy       1426 -----------------------------GTGATAGAGGCGCCTGGTGCCTCAAAATGTA 1456
                                           ||||||||||||||||||||||||||||||||
Db        541 TCACACTCTTGCACTGCATTAGCTGCAAGGTGATAGAGGCGCCTGGTGCCTCAAAATGTA 600

Qy       1457 TTGGCCCC-GGGTGCTGTCACGTGGCGCAG-CCCGACTCGGTGTACTGCA-GTAATGACT 1513
              |||||||| |||||||||||||||||||| |||||||||||||||||||| |||||||||
Db        601 TTGGCCCCGGGGTGCTGTCACGTGGCGCAGCCCCGACTCGGTGTACTGCAGGTAATGACT 660
```

```
Qy       1514 GTATCCTCAAACACGCCGC-AGCGACAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAAG 1572
             |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db        661 GTATCCTCAAACACGCCGCAAGCGACAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAAG 720

Qy       1573 CCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGAAATGCGGT 1632
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGAAATGCGGT 780

Qy       1633 GCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAAAAAAAAGAG 1692
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAAAAAAAGAG 840

Qy       1693 ACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCA 1752
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCA 900

Qy       1753 GCTTGTGAGAGCAGCACG-CCGTCGTGGGCGAG-CGATCACAATTACAATGC-AGTAAAG 1809
             |||||||||||||||||| |||||||||||||| |||||||||||||||||| |||||||
Db        901 GCTTGTGAGAGCAGCACGCCCGTCGTGGGCGAGCCGATCACAATTACAATGCAAGTAAAG 960

Qy       1810 CCAGAAAAGACTGCTGCTCCCTCG-CCGTCACTGTTGTATAAATGTATGTA 1859
             |||||||||||||||||||||||| |||||||||||||||||||  | | |
Db        961 CCAGAAAAGACTGCTGCTCCCTCGCCCGTCACTGTTGTATAAATCCACGAA 1011
```

RESULT 14
AAC08973
ID AAC08973 standard; cDNA; 659 BP.
XX
AC AAC08973;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 13048.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-00200610.
XX
PR 26-FEB-1999; 99US-0122487P.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
PS Claim 1; SEQ ID NO 13048; 71pp + Sequence Listing; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors
XX
SQ Sequence 659 BP; 137 A; 180 C; 135 G; 195 T; 0 U; 12 Other;

Query Match 22.9%; Score 596.8; DB 3; Length 659;
Best Local Similarity 96.8%; Pred. No. 3.4e-150;
Matches 640; Conservative 9; Mismatches 6; Indels 6; Gaps 4;

```
Qy       1844 TGTATAAATGTATGTATCACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGA 1903
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TGTATAAATGTATGTATCACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGA 60

Qy       1904 TAGCCATCCCCTGGGCCTGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACA 1963
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TAGCCATCCCCTGGGCCTGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACA 120

Qy       1964 GACACCGGCTG--CACCATCAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGC 2021
             |||||||||||   :|:||||||||||||||||||||||||||||||||||||||||||:
Db        121 GACACCGGCTGGCAMCMATCAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGY 180

Qy       2022 CCTGTGTCCATCCCTAGTCTGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCT 2081
             :|||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||
Db        181 VCTGTGTCCATCCCTAGTCTGTCAGGACTTCCTGWCACTGTTTTCCAAAGCTGTAAACCT 240

Qy       2082 CACTGGTGAACGTTCACCTTAATGATTGATTC-TTTAATCTCTGTTTTCACTCTCAGGCT 2140
             |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db        241 CACTGGTGAACGTTCACCTTAATGATTGATTCTTTTAATCTCTGTTTTCACTCTCAGGCT 300

Qy       2141 CTGGTAAGTATTTGTATTCTCTTCATCCCAGTCTGATTGCATAGCCACACTGCCCGGCAC 2200
             |||||||||||| ||||:  ||||||||||||||||||||||||||||||||||||||||
Db        301 CTGGTAAGTATTCGTATDNCCTTCATCCCAGTCTGATTGCATAGCCACACTGCCCGGCAC 360

Qy       2201 GCCACATCCACCCCTGTCTGCACATGAGTTGTTCTGACAACAGCGCTGTATACGCTTCAG 2260
             ||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db        361 GCCACATCCACCCCTGTCTGCACATGAGTTKTTCTGACAACAGCGCTGTATACGCTTCAG 420

Qy       2261 TTTTTCCACATTGTCCACGGCCAGCACATGAAAGCATCACTTCTTTTTTATGTTGTGGGA 2320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TTTTTCCACATTGTCCACGGCCAGCACATGAAAGCATCACTTCTTTTTTATGTTGTGGGA 480

Qy       2321 ATCTTTGCAAGTTAGTGTTGCATCTGATTTTCAGGTGTACATTTATTTTTGACTGGGCAG 2380
             ||||||||||||||||:|||  :||||||||||||||||||||||||||||||||||||||
Db        481 ATCTTTGCAAGTTAGTGTKGCANNWGATTTTCAGGTGTACATTTATTTTTGACTGGGCAG 540

Qy       2381 ATAGGGGATTTTTTTTTTTTTCCATGTCCGATTCACACGCTACACACCCACATGAACACA 2440
             |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGGGGA--TTTTTTTTTTTCCATGTCCGATTCACACGCTACACACCCACATGAACACA 598

Qy       2441 TTCGAACTTCGAAGG-CACACACTCCTGCTTCATAGGCCCCACGGTAAGTGAGTTCACAC 2499
             ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        599 TTCGAACTTCGAAGGCCACACACTCCTGCTTCATAGGCCCCACGGTAAGTGAGTTCACAC 658

Qy       2500 C 2500
             |
Db        659 C 659
```

RESULT 15
AAH08005
ID AAH08005 standard; cDNA; 681 BP.
XX
AC AAH08005;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (5'-primer) SEQ ID NO:4840.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.

XX
PD 07-FEB-2001.
XX
PP 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 1; SEQ ID NO 4840; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesising 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 681 BP; 161 A; 177 C; 220 G; 119 T; 0 U; 4 Other;

Query Match 19.6%; Score 510.8; DB 4; Length 681;
Best Local Similarity 98.9%; Pred. No. 5.3e-127;
Matches 534; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

```
Qy        65 GGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTG 124
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       101 GGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTG 160

Qy       125 TTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCC 184
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       161 TTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCC 220

Qy       185 AGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCAAG 244
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       221 AGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCAAG 280

Qy       245 CTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTA 304
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       281 CTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTA 340

Qy       305 AGGCCATCAAACCCACC-AGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACT 363
             ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       341 AGGCCATCAAACCCACCAAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACT 400
```

```
Qy       364 ATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCA 423
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       401 ATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCA 460

Qy       424 CAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGC 483
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       461 CAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGC 520

Qy       484 GTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTG 543
             |||||||| ||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       521 GTGGAGCANTTCCTGACCATTGCGCGGCGCCGCGGCAAGAGGAGCATGCCTGTCTCCCTG 580

Qy       544 GAGGATTCT-GGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGG 602
             ||||||||| |||||||||||||||||||||||||||||||||||||||||||| | ||
Db       581 GAGGATTCTGGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCNAAGG 640
```

Search completed: April 23, 2004, 08:28:12
Job time : 983.654 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2004, 04:42:12 ; Search time 9937.73 Seconds
(without alignments)
11383.418 Million cell updates/sec

Title: US-09-787-016A-1
Perfect score: 2610
Sequence: 1 ctcggtggccgtccgcccac.....actcttaagatcatatcctg 2610

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2610	100.0	2610	6	BD244712	BD244712 Genes enc
2	2610	100.0	2610	6	AX023369	AX023369 Sequence
3	2532.4	97.0	2725	9	BC014489	BC014489 Homo sapi
4	2517.8	96.5	2767	9	BC004237	BC004237 Homo sapi
5	2516.2	96.4	2772	6	AX878224	AX878224 Sequence
6	2516.2	96.4	2772	6	BD157094	BD157094 Primer fo
7	2516.2	96.4	2772	9	AK002127	AK002127 Homo sapi
8	1941	74.4	2407	9	BC000770	BC000770 Homo sapi
9	1023.6	39.2	2602	10	MMU238332	AJ238332 Mus muscu
10	1023.6	39.2	2867	6	BD244713	BD244713 Genes enc
11	1023.6	39.2	2867	6	AX023370	AX023370 Sequence
12	985	37.7	4906	10	AK129117	AK129117 Mus muscu
c 13	948.4	36.3	160241	9	HS885L7	AL035669 Human DNA
14	770.6	29.5	6692	9	AB002331	AB002331 Human mRN
c 15	654.6	25.1	28519	9	HSJ563E14	AL117379 Human DNA
16	596.8	22.9	659	6	AX897185	AX897185 Sequence
17	596.8	22.9	659	6	BD032718	BD032718 Sequence
18	510.8	19.6	681	6	AX869935	AX869935 Sequence
19	510.8	19.6	681	6	BD149997	BD149997 Primer fo
20	491.4	18.8	650	9	HUMYP10D03	AF085876 Homo sapi
21	490.4	18.8	272404	2	AC108337	AC108337 Rattus no
c 22	481.8	18.5	104663	10	AL732560	AL732560 Mouse DNA
23	449.8	17.2	221547	2	AC119716	AC119716 Rattus no
24	386.4	14.8	69252	2	AC101519	AC101519 Mus muscu
c 25	336.6	12.9	542	6	AX874900	AX874900 Sequence
c 26	336.6	12.9	542	6	BD154962	BD154962 Primer fo
27	262.4	10.1	2428	5	BC060442	BC060442 Xenopus l
c 28	232	8.9	69252	2	AC101519	AC101519 Mus muscu
29	202.8	7.8	472	6	AR426988	AR426988 Sequence
30	202.8	7.8	472	6	BD122541	BD122541 EST and e
c 31	103.4	4.0	198133	2	BX296530	BX296530 Danio rer
32	70.8	2.7	2161	3	AK114522	AK114522 Ciona int
c 33	68.6	2.6	61204	2	AC017132	AC017132 Drosophil
34	68.6	2.6	161601	3	AC007594	AC007594 Drosophil
35	68.6	2.6	225655	3	AE003695	AE003695 Drosophil
c 36	67	2.6	7218	6	I66494	I66494 Sequence 14
37	66.2	2.5	125020	9	AF429315	AF429315 Homo sapi
38	65	2.5	2888	9	AK055678	AK055678 Homo sapi
39	65	2.5	5975	6	AX329603	AX329603 Sequence
40	65	2.5	6256	6	AR338778	AR338778 Sequence
41	65	2.5	6936	9	D87685	D87685 Human mRNA
42	65	2.5	6948	6	AX210673	AX210673 Sequence
43	65	2.5	6948	9	AF091622	AF091622 Homo sapi
44	65	2.5	7142	9	HSM808416	BX648268 Homo sapi
c 45	64	2.5	125020	9	AF429315	AF429315 Homo sapi

ALIGNMENTS

RESULT 1
BD244712
LOCUS BD244712 2610 bp DNA linear PAT 17-JUL-2003
DEFINITION Genes encoding for the human and murine death inducer-obliterator-1.
ACCESSION BD244712
VERSION BD244712.1 GI:33054482
KEYWORDS JP 2002526040-A/1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2610)
AUTHORS Alonso,C.M., Domingo,D.G., Grandien,A., Leonardo,E. and Martinez,P.
TITLE Genes encoding for the human and murine death inducer-obliterator-1

```
  JOURNAL   Patent: JP 2002526040-A 1 20-AUG-2002;
            CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002526040-A/1
            PD   20-AUG-2002
            PF   10-SEP-1999 JP 2000570314
            PR   10-SEP-1998 SE    9803069-5,17-SEP-1998 US    60/100873 PI
            CARLOS MARTINEZ ALONSO,DAVID GARCIA DOMINGO,ALF GRANDIEN, PI
            ESTHER LEONARDO,
            PI   PEDRO MARTINEZ
            PC   C12N15/09,A61K38/00,A61K45/00,A61K48/00,A61P3/10,A61P17/00, PC
              A61P29/00,
            PC   A61P35/00,A61P37/06,A61P43/00,C07K14/47,C07K16/18,C12N1/21, PC
              C12N5/10,
            PC   C12P21/02,C12Q1/02,C12Q1/68,G01N33/15,G01N33/50//C12P21/08, PC
              C12N15/00,
            PC   C12N5/00,A61K37/02
            CC   This gene is referred to in the application as Human Death  CC
                   Inducer
            CC   Obliterator Gene 1.
            CC   It has now been named by the Human Gene
            Nomenclature Committee
            CC             as
            CC   Death-Associated Transcription Factor (DATF-1) FH   Key
                 Location/Qualifiers
            FT   source          1. .2610
            FT                   /organism='Homo sapiens (human)'.
FEATURES             Location/Qualifiers
     source          1. .2610
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             100.0%;  Score 2610;  DB 6;  Length 2610;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2610;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTCGGTGGCCGTCCGCCCACTCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTCGGTGGCCGTCCGCCCACTCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGC 60

Qy         61 CTGCGGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTGCGGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAG 120

Qy        121 CCTGTTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCTGTTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGG 180

Qy        181 ACCCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTC 240

Qy        241 CAAGCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAAGCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCA 300

Qy        301 CCTAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCTAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACC 360

Qy        361 ACTATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACTATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCC 420

Qy        421 CCACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAG 480

Qy        481 CGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCC 540

Qy        541 CTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAG 600

Qy        601 GGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTG 660

Qy        661 AAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCC 720

Qy        721 GATAGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GATAGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGG 780

Qy        781 GAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGC 840

Qy        841 CGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTC 900

Qy        901 CTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGAT 960

Qy        961 GACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGA 1020

Qy       1021 GACTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GACTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATT 1080

Qy       1081 TGCCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGCCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTT 1140

Qy       1141 CATGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CATGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAA 1200

Qy       1201 GACTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GACTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACG 1260

Qy       1261 GCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGT 1320

Qy       1321 ATAGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ATAGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAA 1380

Qy       1381 GCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCT 1440

Qy       1441 GGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTAC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTAC 1500

Qy       1501 TGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGT 1560

Qy       1561 AAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 AAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTT 1620
```

```
Qy       1621 CCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCA 1680
Db       1621 CCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCA 1680

Qy       1681 GAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTC 1740
Db       1681 GAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTC 1740

Qy       1741 GGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAAT 1800
Db       1741 GGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAAT 1800

Qy       1801 GCAGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTAT 1860
Db       1801 GCAGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTAT 1860

Qy       1861 CACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCC 1920
Db       1861 CACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCC 1920

Qy       1921 TGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCAT 1980
Db       1921 TGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCAT 1980

Qy       1981 CAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTC 2040
Db       1981 CAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTC 2040

Qy       2041 TGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGTGAACGTTCACCT 2100
Db       2041 TGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGTGAACGTTCACCT 2100

Qy       2101 TAATGATTGATTCTTTAATCTCTGTTTTCACTCTCAGGCTCTGGTAAGTATTTGTATTCT 2160
Db       2101 TAATGATTGATTCTTTAATCTCTGTTTTCACTCTCAGGCTCTGGTAAGTATTTGTATTCT 2160

Qy       2161 CTTCATCCCAGTCTGATTGCATAGCCACACTGCCCGGCACGCCACATCCACCCCTGTCTG 2220
Db       2161 CTTCATCCCAGTCTGATTGCATAGCCACACTGCCCGGCACGCCACATCCACCCCTGTCTG 2220

Qy       2221 CACATGAGTTGTTCTGACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGG 2280
Db       2221 CACATGAGTTGTTCTGACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGG 2280

Qy       2281 CCAGCACATGAAAGCATCACTTCTTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTG 2340
Db       2281 CCAGCACATGAAAGCATCACTTCTTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTG 2340

Qy       2341 CATCTGATTTTCAGGTGTACATTTATTTTTGACTGGGCAGATAGGGGATTTTTTTTTTTT 2400
Db       2341 CATCTGATTTTCAGGTGTACATTTATTTTTGACTGGGCAGATAGGGGATTTTTTTTTTTT 2400

Qy       2401 TCCATGTCCGATTCACACGCTACACACCCACATGAACACATTCGAACTTCGAAGGCACAC 2460
Db       2401 TCCATGTCCGATTCACACGCTACACACCCACATGAACACATTCGAACTTCGAAGGCACAC 2460

Qy       2461 ACTCCTGCTTCATAGGCCCCACGGTAAGTGAGTTCACACCTAGAACACTGTCCTGACCGC 2520
Db       2461 ACTCCTGCTTCATAGGCCCCACGGTAAGTGAGTTCACACCTAGAACACTGTCCTGACCGC 2520

Qy       2521 AGGACGCGTGCCTTGGACTTGGTATTCTACATGTGACTGGCTTTCTTGCCCTCGTCTCTT 2580
Db       2521 AGGACGCGTGCCTTGGACTTGGTATTCTACATGTGACTGGCTTTCTTGCCCTCGTCTCTT 2580

Qy       2581 GAATGTTTAGACTCTTAAGATCATATCCTG 2610
Db       2581 GAATGTTTAGACTCTTAAGATCATATCCTG 2610
```

RESULT 2
AX023369
LOCUS AX023369 2610 bp DNA linear PAT 15-SEP-2000
DEFINITION Sequence 1 from Patent WO0015787.

ACCESSION AX023369
VERSION AX023369.1 GI:10183781
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Leonardo,E., Martinez,P., Alonso,C.M., Domingo,D.G. and Grandien,A.
TITLE Genes encoding for the human and murine death in ducer-obliterator-1
JOURNAL Patent: WO 0015787-A 1 23-MAR-2000;
LEONARDO ESTHER (ES) ; MARTINEZ PEDRO (ES) ; CONSEJO SUPERIOR INVESTIGACION (ES) ; DOMINGO DAVID GARCIA (ES) ; MARTINEZ ALONSO CARLOS (ES) ; BANNERMAN DAVID GARDNER (GB) ; GRANDIEN ALF (SE)
FEATURES Location/Qualifiers
source 1. .2610
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="This gene is referred to in the application as Human Death Inducer Obliterator Gene 1. It has now been named by the Human Gene Nomenclature Committee as Death-Associated Transcription Factor (DATF-1)"
ORIGIN

Query Match 100.0%; Score 2610; DB 6; Length 2610;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2610; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CTCGGTGGCCGTCCGCCCACTCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGC 60
Db          1 CTCGGTGGCCGTCCGCCCACTCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGC 60

Qy         61 CTGCGGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAG 120
Db         61 CTGCGGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAG 120

Qy        121 CCTGTTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGG 180
Db        121 CCTGTTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGG 180

Qy        181 ACCCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTC 240
Db        181 ACCCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTC 240

Qy        241 CAAGCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCA 300
Db        241 CAAGCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCA 300

Qy        301 CCTAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACC 360
Db        301 CCTAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACC 360

Qy        361 ACTATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCC 420
Db        361 ACTATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCC 420

Qy        421 CCACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAG 480
Db        421 CCACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAG 480

Qy        481 CGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCC 540
Db        481 CGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCC 540

Qy        541 CTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAG 600
Db        541 CTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAG 600

Qy        601 GGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTG 660
Db        601 GGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTG 660
```

```
Qy      661 AAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCC 720
Db      661 AAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCC 720

Qy      721 GATAGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGG 780
Db      721 GATAGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGG 780

Qy      781 GAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGC 840
Db      781 GAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGC 840

Qy      841 CGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTC 900
Db      841 CGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTC 900

Qy      901 CTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGAT 960
Db      901 CTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGAT 960

Qy      961 GACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGA 1020
Db      961 GACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGA 1020

Qy     1021 GACTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATT 1080
Db     1021 GACTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATT 1080

Qy     1081 TGCCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTT 1140
Db     1081 TGCCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTT 1140

Qy     1141 CATGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAA 1200
Db     1141 CATGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAA 1200

Qy     1201 GACTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACG 1260
Db     1201 GACTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACG 1260

Qy     1261 GCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGT 1320
Db     1261 GCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGT 1320

Qy     1321 ATAGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAA 1380
Db     1321 ATAGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAA 1380

Qy     1381 GCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCT 1440
Db     1381 GCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCT 1440

Qy     1441 GGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTAC 1500
Db     1441 GGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTAC 1500

Qy     1501 TGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGT 1560
Db     1501 TGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGT 1560

Qy     1561 AAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTT 1620
Db     1561 AAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTT 1620

Qy     1621 CCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCA 1680
Db     1621 CCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCA 1680

Qy     1681 GAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTC 1740
Db     1681 GAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTC 1740

Qy     1741 GGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAAT 1800
Db     1741 GGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAAT 1800

Qy     1801 GCAGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTAT 1860
Db     1801 GCAGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTAT 1860

Qy     1861 CACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCC 1920
Db     1861 CACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCC 1920

Qy     1921 TGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCAT 1980
Db     1921 TGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCAT 1980

Qy     1981 CAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTC 2040
Db     1981 CAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTC 2040

Qy     2041 TGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGTGAACGTTCACCT 2100
Db     2041 TGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGTGAACGTTCACCT 2100

Qy     2101 TAATGATTGATTCTTTAATCTCTGTTTTCACTCTCAGGCTCTGGTAAGTATTTGTATTCT 2160
Db     2101 TAATGATTGATTCTTTAATCTCTGTTTTCACTCTCAGGCTCTGGTAAGTATTTGTATTCT 2160

Qy     2161 CTTCATCCCAGTCTGATTGCATAGCCACACTGCCCGGCACGCCACATCCACCCCTGTCTG 2220
Db     2161 CTTCATCCCAGTCTGATTGCATAGCCACACTGCCCGGCACGCCACATCCACCCCTGTCTG 2220

Qy     2221 CACATGAGTTGTTCTGACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGG 2280
Db     2221 CACATGAGTTGTTCTGACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGG 2280

Qy     2281 CCAGCACATGAAAGCATCACTTCTTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTG 2340
Db     2281 CCAGCACATGAAAGCATCACTTCTTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTG 2340

Qy     2341 CATCTGATTTTCAGGTGTACATTTATTTTTGACTGGGCAGATAGGGGATTTTTTTTTTTT 2400
Db     2341 CATCTGATTTTCAGGTGTACATTTATTTTTGACTGGGCAGATAGGGGATTTTTTTTTTTT 2400

Qy     2401 TCCATGTCCGATTCACACGCTACACACCCACATGAACACATTCGAACTTCGAAGGCACAC 2460
Db     2401 TCCATGTCCGATTCACACGCTACACACCCACATGAACACATTCGAACTTCGAAGGCACAC 2460

Qy     2461 ACTCCTGCTTCATAGGCCCCACGGTAAGTGAGTTCACACCTAGAACACTGTCCTGACCGC 2520
Db     2461 ACTCCTGCTTCATAGGCCCCACGGTAAGTGAGTTCACACCTAGAACACTGTCCTGACCGC 2520

Qy     2521 AGGACGCGTGCCTTGGACTTGGTATTCTACATGTGACTGGCTTTCTTGCCCTCGTCTCTT 2580
Db     2521 AGGACGCGTGCCTTGGACTTGGTATTCTACATGTGACTGGCTTTCTTGCCCTCGTCTCTT 2580

Qy     2581 GAATGTTTAGACTCTTAAGATCATATCCTG 2610
Db     2581 GAATGTTTAGACTCTTAAGATCATATCCTG 2610
```

RESULT 3
BC014489

```
LOCUS       BC014489               2725 bp    mRNA    linear   PRI 04-OCT-2003
DEFINITION  Homo sapiens death associated transcription factor 1, transcript
            variant 2, mRNA (cDNA clone MGC:23216 IMAGE:4901057), complete cds.
ACCESSION   BC014489
VERSION     BC014489.1  GI:15680266
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 2725)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
    TITLE   Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2725)
  AUTHORS   Strausberg,R.
    TITLE   Direct Submission
  JOURNAL   Submitted (17-SEP-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
   REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 34 Row: n Column: 1
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Similarity but not
            identity to protein.
FEATURES             Location/Qualifiers
     source          1. .2725
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:23216 IMAGE:4901057"
                     /tissue_type="Muscle, rhabdomyosarcoma"
                     /clone_lib="NIH_MGC_17"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     gene            1. .2725
                     /gene="DATF1"
                     /note="synonyms: DIO1, FLJ11265, KIAA0333"
                     /db_xref="LocusID:11083"
                     /db_xref="MIM:604140"
     CDS             238. .1926
                     /codon_start=1
                     /product="death associated transcription factor 1, isoform
                     a"
                     /protein_id="AAH14489.1"
                     /db_xref="GI:15680267"
                     /db_xref="LocusID:11083"
                     /translation="MDDKGDPSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDAE
                     ADPLEPPPPPQQQLGLSLRRSGRQPKRTERVEQFLTIARRRGRRSMPVSLEDSGEPTSC
                     PATDAETASEGSVESASETRSGPQSASTAVKERPASSEKVKGGDDHDDTSDSDSDGLT
                     LKELQNRLRRKREQEPTERPLKGIQSRLRKKRREEGPAETVGSEASDTVEGVLPSKQE
                     PENDQGVVSQAGKDDRESKLEGKAAQDIKDEEPGDLGRPKPECEGYDPNALYCICRQP
                     HNNRPMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETHSETAD
                     QQEAKWRPGDADGTDCTSIGTIEQKSSEDQGIKGRIEKAANPSGKKKLKIFQPVIEAP
                     GASKCIGPGCCHVAQPDSVYCSNDCILKHAAATMKFLSSGKEQKPKPKEKMKMKPEKP
                     SLPKCGAQAGIKISSVHKRPAPEKKETTVKKAVVVPARSEALGKEAACESSTPSWASD
                     HNYNAVKPEKTAAPSPSLLYKCMYHLGVGLLDPSRSFWIAIPWACPGLGVAALC"
     misc_feature    1048. .1197
                     /note="PHD; Region: PHD-finger. PHD folds into an
                     interleaved type of Zn-finger chelating 2 Zn ions in a
                     similar manner to that of the RING and FYVE domains"
                     /db_xref="CDD:pfam00628"
ORIGIN

  Query Match             97.0%;  Score 2532.4;  DB 9;  Length 2725;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 2556;  Conservative    0;  Mismatches    7;  Indels    3;  Gaps    2;

Qy         46 CGAGACCCTAGAGGCCTGCGGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTT 105
               | ||| | | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         19 CTAGAGGCCTGCGGCCTGCGGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTT 78

Qy        106 GTTGAAAATACGAAGCCTGTTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCC 165
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         79 GTTGAAAATACGAAGCCTGTTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCC 138

Qy        166 TGGCTGTCTGCTTGGACCCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATT 225
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 TGGCTGTCTGCTTGGACCCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATT 198

Qy        226 TCAGGGAAAAGTGTCCAAGCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCG 285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 TCAGGGAAAAGTGTCCAAGCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCG 258

Qy        286 AGCAATGAGGAGGCACCTAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGG 345
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 AGCAATGAGGAGGCACCTAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGG 318

Qy        346 GGTTTTCGAAGGACCACTATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCA 405
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 GGTTTTCGAAGGACCACTATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCA 378

Qy        406 CTGGAGCCGCCACCCCCACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAG 465
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 CTGGAGCCGCCACCCCCACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAG 438

Qy        466 CCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGG 525
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        439 CCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGG 498

Qy        526 AGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCC 585
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 AGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCC 558

Qy        586 GAGACAGCCTCCGAGGGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCT 645
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        559 GAGACAGCCTCCGAGGGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCT 618

Qy        646 GCTTCCACAGCTGTGAAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGAC 705
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        619 GCTTCCACAGCTGTGAAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGAC 678

Qy        706 CACGATGACACCTCCGATAGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGC 765
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        679 CACGATGACACCTCCGATAGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGC 738
```

```
Qy       766 CTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGC 825
Db       739 CTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGC 798

Qy       826 CTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGAC 885
Db       799 CTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGAC 858

Qy       886 ACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCC 945
Db       859 ACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCC 918

Qy       946 CAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAA 1005
Db       919 CAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAA 978

Qy      1006 GATGAGGAGCCTGGAGACTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAAC 1065
Db       979 GATGAGGAGCCTGGAGACTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAAC 1038

Qy      1066 GCCCTGTATTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGC 1125
Db      1039 GCCCTGTATTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGC 1098

Qy      1126 TGTGAAGAATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTG 1185
Db      1099 TGTGAAGAATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTG 1158

Qy      1186 GAAAGGAATGGGGAAGACTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAG 1245
Db      1159 GAAAGGAATGGGGAAGACTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAG 1218

Qy      1246 ACTCATTCAGAAACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGC 1305
Db      1219 ACTCATTCAGAAACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGC 1278

Qy      1306 ACCGATTGTACAAGTATAGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAG 1365
Db      1279 ACCGATTGTACAAGTATAGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAG 1338

Qy      1366 GGTAGAATTGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCT 1425
Db      1339 GGTAGAATTGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCT 1398

Qy      1426 GTGATAGAGGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAG 1485
Db      1399 GTGATAGAGGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAG 1458

Qy      1486 CCCGACTCGGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAG 1545
Db      1459 CCCGACTCGGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAG 1518

Qy      1546 TTTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCA 1605
Db      1519 TTTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCA 1578

Qy      1606 GAGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCAC 1665
Db      1579 GAGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCAC 1638

Qy      1666 AAGAGACCAGCTCCAGAAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCG 1725
Db      1639 AAGAGACCAGCTCCAGAAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCG 1698

Qy      1726 CGGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGC 1785
Db      1699 CGGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGC 1758

Qy      1786 GATCACAATTACAATGCAGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTG 1845
Db      1759 GATCACAATTACAATGCAGTAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTG 1818

Qy      1846 TATAAATGTATGTATCACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATA 1905
Db      1819 TATAAATGTATGTATCACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATA 1878

Qy      1906 GCCATCCCCTGGGCCTGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGA 1965
Db      1879 GCCATCCCCTGGGCCTGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGA 1938

Qy      1966 CACCGGCTGCACCATCAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTG 2025
Db      1939 CACCGGCTGCACCATCAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTG 1998

Qy      2026 TGTCCATCCCTAGTCTGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACT 2085
Db      1999 TGTCCATCCCTAGTCTGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACT 2058

Qy      2086 GGTGAACGTTCACCTTAATGATTGATTCTTTAATCTCTGTTTTCACTCTCAGGCTCTGGT 2145
Db      2059 GGTGAACGTTCACCTTAATGATTGATTCTTTAATCTCTGTTTTCACTCTCAGGCTCTGGT 2118

Qy      2146 AAGTATTTGTATTCTCTTCATCCCAGTCTGATTGCATAGCCACACTGCCCGGCACGCCAC 2205
Db      2119 AAGTATTCGTATTCTCTTCATCCCAGTCTGATTGCATAGCCACACTGCCCGGCACGCCAC 2178

Qy      2206 ATCCACCCCTGTCTGCACATGAGTTGTTCTGACAACAGCGCTGTATACGCTTCAGTTTTT 2265
Db      2179 ATCCACCCCTGTCTGCACATGAGTTGTTCTGACAACAGCGCTGTATACGCTTCAGTTTTT 2238

Qy      2266 CCACATTGTCCACGGCCAGCACATGAAAGCATCACTTCTTTTTTATGTTGTGGGAATCTT 2325
Db      2239 CCACATTGTCCACGGCCAGCACATGAAAGCATCACTTCTTTTTTATGTTGTGGGAATCTT 2298

Qy      2326 TGCAAGTTAGTGTTGCATCTGATTTTCAGGTGTACATTTATTTTTGACTGGGCAGATAGG 2385
Db      2299 TGCAAGTTAGTGTTGCATCTGATTTTCAGGTGTACATTTATTTTTGACTGGGCAGATAGG 2358

Qy      2386 GGATTTTTTTTTTTTTCCATGTCCGATTCACACGCTACACACCCACATGAACACATTCGA 2445
Db      2359 GGA--TTTTTTTTTTCCATGTCCGATTCACACGCTACACACCCACATGAACACATTCGA 2416

Qy      2446 ACTTCGAAGG-CACACACTCCTGCTTCATAGGCCCCACGGTAAGTGAGTTCACACCTAGA 2504
Db      2417 ACTTCGAAGGCCACACACTCCTGCTTCATAGGCCCCACGGTAAGTGAGTTCACACCTAGA 2476

Qy      2505 ACACTGTCCTGACCGCAGGACGCGTGCCTTGGACTTGGTATTCTACATGTGACTGGCTTT 2564
Db      2477 ACACTGTCCTGACCGCAGGACGCGTGCCTTGGACTTGGTATTCTACATGTGACTGGCTTT 2536

Qy      2565 CTTGCCCTCGTCTCTTGAATGTTTAGACTCTTAAGATCATATCCTG 2610
Db      2537 CTTGCCCTCGTCTCTTGAATGTTTAGACTCTTAAGATCATATCCTG 2582
```

RESULT 4
BC004237

```
LOCUS       BC004237               2767 bp    mRNA    linear   PRI 12-NOV-2003
DEFINITION  Homo sapiens death associated transcription factor 1, transcript
            variant 1, mRNA (cDNA clone MGC:896 IMAGE:2959493), complete cds.
ACCESSION   BC004237
VERSION     BC004237.2  GI:33872756
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2767)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
```

```
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2767)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     On Aug 19, 2003 this sequence version replaced gi:13278980.
            Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 2 Row: o Column: 16
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 18375620.
FEATURES            Location/Qualifiers
     source         1. .2767
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="MGC:896 IMAGE:2959493"
                    /tissue_type="Colon, adenocarcinoma"
                    /clone_lib="NIH_MGC_15"
                    /lab_host="DH10B-R"
                    /note="Vector: pOTB7"
     gene           1. .2767
                    /gene="DATF1"
                    /note="synonyms: DIO1, FLJ11265, KIAA0333"
                    /db_xref="LocusID:11083"
                    /db_xref="MIM:604140"
     CDS            282. .1970
                    /codon_start=1
                    /product="death associated transcription factor 1, isoform
                    a"
                    /protein_id="AAH04237.1"
                    /db_xref="GI:13278981"
                    /db_xref="LocusID:11083"
                    /translation="MDDKGDPSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDAE
                    ADPLEPPPPQQQLGLSLRRSGRQPKRTERVEQFLTIARRRGRRSMPVSLEDSGEPTSC
                    PATDAETASEGSVESASETRSGPQSASTAVKERPASSEKVKGGDDHDDTSDSDSDGLT
                    LKELQNRLRRKREQEPTERPLKGIQSRLRKKRREEGPAETVGSEASDTVEGVLPSKQE
                    PENDQGVVSQAGKDDRESKLEGKAAQDIKDEEPGDLGRPKPECEGYDPNALYCICRQP
                    HNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETHSETAD
                    QQEAKWRPGDADGTDCTSIGTIEQKSSEDQGIKGRIEKAANPSGKKKLKIFQPVIEAP
                    GASKCIGPGCCHVAQPDSVYCSNDCILKHAAATMKFLSSGKEQKPKPKEKMKMKPEKP
                    SLPKCGAQAGIKISSVHKRPAPEKKETTVKKAVVVPARSEALGKEAACESSTPSWASD
                    HNYNAVKPEKTAAPSPSLLYKCMYHLGVGLLDPSRSFWIAIPWACPGLGVAALC"
     misc_feature   1092. .1241
                    /note="PHD; Region: PHD-finger. PHD folds into an
                    interleaved type of Zn-finger chelating 2 Zn ions in a
                    similar manner to that of the RING and FYVE domains"
                    /db_xref="CDD:pfam00628"
ORIGIN

  Query Match             96.5%;  Score 2517.8;  DB 9;  Length 2767;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 2542;  Conservative    0;  Mismatches    2;  Indels    3;  Gaps    2;

Qy         65 GGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTG 124
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         82 GGAGCTTATTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTG 141

Qy        125 TTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCC 184
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        142 TTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCC 201

Qy        185 AGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCAAG 244
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        202 AGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCAAG 261

Qy        245 CTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTA 304
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        262 CTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTA 321

Qy        305 AGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTA 364
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        322 AGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTA 381

Qy        365 TCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCAC 424
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        382 TCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCAC 441

Qy        425 AGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCG 484
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        442 AGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCG 501

Qy        485 TGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGG 544
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        502 TGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGG 561

Qy        545 AGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCA 604
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        562 AGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCA 621

Qy        605 GCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGG 664
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        622 GCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGG 681

Qy        665 AACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATA 724
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        682 AACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATA 741

Qy        725 GTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAAC 784
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        742 GTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAAC 801

Qy        785 AGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGG 844
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        802 AGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGG 861

Qy        845 AGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGC 904
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        862 AGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGC 921

Qy        905 CCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACA 964
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        922 CCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACA 981

Qy        965 GAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACT 1024
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        982 GAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACT 1041
```

```
Qy       1025 TGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCC 1084
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1042 TGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCC 1101

Qy       1085 GCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATG 1144
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1102 GCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATG 1161

Qy       1145 GCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACT 1204
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1162 GCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACT 1221

Qy       1205 ATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAG 1264
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1222 ATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAG 1281

Qy       1265 ATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAG 1324
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1282 ATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAG 1341

Qy       1325 GAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTG 1384
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1342 GAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTG 1401

Qy       1385 CAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGGTG 1444
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1402 CAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGGTG 1461

Qy       1445 CCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTGCA 1504
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1462 CCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTGCA 1521

Qy       1505 GTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAAAG 1564
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1522 GTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAAAG 1581

Qy       1565 AACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGA 1624
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1582 AACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGA 1641

Qy       1625 AATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAA 1684
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1642 AATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAA 1701

Qy       1685 AAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGA 1744
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1702 AAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGA 1761

Qy       1745 AGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAG 1804
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1762 AGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAG 1821

Qy       1805 TAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACC 1864
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1822 TAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACC 1881

Qy       1865 TAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCCTGTC 1924
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1882 TAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCCTGTC 1941

Qy       1925 CAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCAGC 1984
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1942 CAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCAGC 2001

Qy       1985 GGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTCTGTC 2044
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2002 GGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTCTGTC 2061

Qy       2045 AGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGTGAACGTTCACCTTAAT 2104
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2062 AGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGTGAACGTTCACCTTAAT 2121

Qy       2105 GATTGATTCTTTAATCTCTGTTTTCACTCTCAGGCTCTGGTAAGTATTTGTATTCTCTTC 2164
             ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       2122 GATTGATTCTTTAATCTCTGTTTTCACTCTCAGGCTCTGGTAAGTATTCGTATTCTCTTC 2181

Qy       2165 ATCCCAGTCTGATTGCATAGCCACACTGCCCGGCACGCCACATCCACCCCTGTCTGCACA 2224
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2182 ATCCCAGTCTGATTGCATAGCCACACTGCCCGGCACGCCACATCCACCCCTGTCTGCACA 2241

Qy       2225 TGAGTTGTTCTGACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGGCCAG 2284
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2242 TGAGTTGTTCTGACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGGCCAG 2301

Qy       2285 CACATGAAAGCATCACTTCTTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTGCATC 2344
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2302 CACATGAAAGCATCACTTCTTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTGCATC 2361

Qy       2345 TGATTTTCAGGTGTACATTTATTTTTGACTGGGCAGATAGGGGATTTTTTTTTTTTTCCA 2404
             ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Db       2362 TGATTTTCAGGTGTACATTTATTTTTGACTGGGCAGATAGGGGA--TTTTTTTTTTTCCA 2419

Qy       2405 TGTCCGATTCACACGCTACACACCCACATGAACACATTCGAACTTCGAAGG-CACACACT 2463
             ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db       2420 TGTCCGATTCACACGCTACACACCCACATGAACACATTCGAACTTCGAAGGCCACACACT 2479

Qy       2464 CCTGCTTCATAGGCCCCACGGTAAGTGAGTTCACACCTAGAACACTGTCCTGACCGCAGG 2523
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2480 CCTGCTTCATAGGCCCCACGGTAAGTGAGTTCACACCTAGAACACTGTCCTGACCGCAGG 2539

Qy       2524 ACGCGTGCCTTGGACTTGGTATTCTACATGTGACTGGCTTTCTTGCCCTCGTCTCTTGAA 2583
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2540 ACGCGTGCCTTGGACTTGGTATTCTACATGTGACTGGCTTTCTTGCCCTCGTCTCTTGAA 2599

Qy       2584 TGTTTAGACTCTTAAGATCATATCCTG 2610
             |||||||||||||||||||||||||||
Db       2600 TGTTTAGACTCTTAAGATCATATCCTG 2626
```

```
RESULT 5
AX878224
LOCUS       AX878224               2772 bp    DNA     linear   PAT 17-DEC-2003
DEFINITION  Sequence 13129 from Patent EP1074617.
ACCESSION   AX878224
VERSION     AX878224.1  GI:40032960
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
            Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
  TITLE     Primers for synthesising full-length cDNA and their use
  JOURNAL   Patent: EP 1074617-A 13129 07-FEB-2001;
            Research Association for Biotechnology (JP)
FEATURES             Location/Qualifiers
     source          1. .2772
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     CDS             301. .1989
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAE90061.1"
                     /db_xref="GI:40032961"
                     /translation="MDDKGDPSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDAE
                     ADPLEPPPPQQQLGLSLRRSGRQPKRTERVEQFLTIARRRGRRSMPVSLEDSGEPTSC
                     PATDAETASEGSVESASETRSGPQSASTAVKERPASSEKVKGGDDHDDTSDSDSDGLT
                     LKELQNRLRRKREQEPTERPLKGIQSRLRKKRREEGPAETVGSEASDTVEGVLPSKQE
                     PENDQGVVSQAGKDDRESKLEGKAAQDIKDEEPEDLGRPKPECEGYDPNALYCICRQP
                     HNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETHSETAD
                     QQEAKWRPGDADGTDCTSIGTIEQKSSEDQGIKGRIEKAANPSGKKKLKIFQPVIEAP
                     GASKCIGPGCCHVAQPDSVYCSNDCILKHAAATMKPLSSGKEQKPKPKEKMKMKPEKP
                     SLPKCGAQAGIKISSVHKRPAPEKKETTVKKAVVVPARSEALGKEAACESSTPSWASD
```

HNYNAVKPEKTAAPSPSLLYKCMYHLGVGLLDPSRSFWIAIPWACPGLGVAALC"

ORIGIN

Query Match 96.4%; Score 2516.2; DB 6; Length 2772;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2541; Conservative 0; Mismatches 3; Indels 3; Gaps 2;

```
Qy      65 GGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTG 124
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     101 GGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTG 160

Qy     125 TTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCC 184
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     161 TTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCC 220

Qy     185 AGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCAAG 244
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     221 AGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCAAG 280

Qy     245 CTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTA 304
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     281 CTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTA 340

Qy     305 AGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTA 364
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     341 AGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTA 400

Qy     365 TCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCAC 424
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     401 TCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCAC 460

Qy     425 AGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCG 484
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     461 AGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCG 520

Qy     485 TGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGG 544
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     521 TGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGG 580

Qy     545 AGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCA 604
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     581 AGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCA 640

Qy     605 GCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGG 664
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     641 GCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGG 700

Qy     665 AACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATA 724
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     701 AACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATA 760

Qy     725 GTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAAC 784
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     761 GTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAAC 820

Qy     785 AGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGG 844
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     821 AGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGG 880

Qy     845 AGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGC 904
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     881 AGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGC 940

Qy     905 CCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACA 964
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     941 CCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACA 1000

Qy     965 GAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACT 1024
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1001 GAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGAAGACT 1060

Qy    1025 TGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCC 1084
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    1061 TGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCC 1120

Qy    1085 GCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATG 1144
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1121 GCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATG 1180

Qy    1145 GCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACT 1204
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1181 GCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACT 1240

Qy    1205 ATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAG 1264
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1241 ATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAG 1300

Qy    1265 ATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAG 1324
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1301 ATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAG 1360

Qy    1325 GAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTG 1384
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1361 GAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTG 1420

Qy    1385 CAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGGTG 1444
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1421 CAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGGTG 1480

Qy    1445 CCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTGCA 1504
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1481 CCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTGCA 1540

Qy    1505 GTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAAAG 1564
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1541 GTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAAAG 1600

Qy    1565 AACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGA 1624
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1601 AACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGA 1660

Qy    1625 AATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAA 1684
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1661 AATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAA 1720

Qy    1685 AAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGA 1744
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1721 AAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGA 1780

Qy    1745 AGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAG 1804
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1781 AGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAG 1840

Qy    1805 TAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACC 1864
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1841 TAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACC 1900

Qy    1865 TAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCCTGGGCCTGTC 1924
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1901 TAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCCTGGGCCTGTC 1960

Qy    1925 CAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCAGC 1984
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1961 CAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCAGC 2020

Qy    1985 GGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTCTGTC 2044
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2021 GGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTCTGTC 2080

Qy    2045 AGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGTGAACGTTCACCTTAAT 2104
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2081 AGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGTGAACGTTCACCTTAAT 2140

Qy    2105 GATTGATTCTTTAATCTCTGTTTTCACTCTCAGGCTCTGGTAAGTATTTGTATTCTCTTC 2164
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2141 GATTGATTCTTTAATCTCTGTTGTCACTCTCAGGCTCTGGTAAGTATTCGTATTCTCTTC 2200
```

```
Qy       2165 ATCCCAGTCTGATTGCATAGCCACACTGCCCGGCACGCCACATCCACCCCTGTCTGCACA 2224
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2201 ATCCCAGTCTGATTGCATAGCCACACTGCCCGGCACGCCACATCCACCCCTGTCTGCACA 2260

Qy       2225 TGAGTTGTTCTGACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGGCCAG 2284
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2261 TGAGTTGTTCTGACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGGCCAG 2320

Qy       2285 CACATGAAAGCATCACTTCTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTGCATC 2344
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2321 CACATGAAAGCATCACTTCTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTGCATC 2380

Qy       2345 TGATTTTCAGGTGTACATTTATTTTTGACTGGGCAGATAGGGGATTTTTTTTTTTTCCA 2404
              ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||
Db       2381 TGATTTTCAGGTGTACATTTATTTTTGACTGGCAGATAGGGGA--TTTTTTTTTTTCCA 2438

Qy       2405 TGTCCGATTCACACGCTACACACCCACATGAACACATTCGAACTTCGAAGG-CACACACT 2463
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db       2439 TGTCCGATTCACACGCTACACACCCACATGAACACATTCGAACTTCGAAGGCCACACACT 2498

Qy       2464 CCTGCTTCATAGGCCCCACGGTAAGTGAGTTCACACCTAGAACACTGTCCTGACCGCAGG 2523
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2499 CCTGCTTCATAGGCCCCACGGTAAGTGAGTTCACACCTAGAACACTGTCCTGACCGCAGG 2558

Qy       2524 ACGCGTGCCTTGGACTTGGTATTCTACATGTGACTGGCTTTCTTGCCCTCGTCTCTTGAA 2583
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2559 ACGCGTGCCTTGGACTTGGTATTCTACATGTGACTGGCTTTCTTGCCCTCGTCTCTTGAA 2618

Qy       2584 TGTTTAGACTCTTAAGATCATATCCTG 2610
              |||||||||||||||||||||||||||
Db       2619 TGTTTAGACTCTTAAGATCATATCCTG 2645
```

```
RESULT 6
BD157094
LOCUS       BD157094                2772 bp    DNA     linear   PAT 17-JAN-2003
DEFINITION  Primer for synthesizing full-length cDNA and use thereof.
ACCESSION   BD157094
VERSION     BD157094.1  GI:27862852
KEYWORDS    JP 2002191363-A/11937.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2772)
  AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
            Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
  TITLE     Primer for synthesizing full-length cDNA and use thereof
  JOURNAL   Patent: JP 2002191363-A 11937 09-JUL-2002;
            HELIX RESEARCH INSTITUTE
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002191363-A/11937
            PD   09-JUL-2002
            PF   28-JUL-2000 JP 2000280990
            PI   TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
            PI   SAITO,
            PI   JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
            PI   KEIICHI NAGAI,TETSUJI OTSUKI
            PC
            C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
             10,
            PC   C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
             Primer for synthesizing full-length cDNA and use thereof FH    Key
                       Location/Qualifiers
            FT   CDS             (301)..(1986).
FEATURES             Location/Qualifiers
     source          1..2772
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN
```

```
  Query Match             96.4%;  Score 2516.2;  DB 6;  Length 2772;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 2541;  Conservative    0;  Mismatches    3;  Indels    3;  Gaps    2;

Qy         65 GGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTG 124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTG 160

Qy        125 TTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCC 184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 TTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCC 220

Qy        185 AGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCAAG 244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCAAG 280

Qy        245 CTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTA 304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 CTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTA 340

Qy        305 AGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTA 364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 AGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTA 400

Qy        365 TCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCAC 424
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 TCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCAC 460

Qy        425 AGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCG 484
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 AGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCG 520

Qy        485 TGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGG 544
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 TGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGG 580

Qy        545 AGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCA 604
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 AGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCA 640

Qy        605 GCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGG 664
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 GCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGG 700

Qy        665 AACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATA 724
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 AACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATA 760

Qy        725 GTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAAC 784
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 GTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAAC 820

Qy        785 AGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGG 844
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 AGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGG 880

Qy        845 AGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGC 904
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 AGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGC 940

Qy        905 CCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACA 964
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 CCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACA 1000

Qy        965 GAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACT 1024
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1001 GAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGAAGACT 1060

Qy       1025 TGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCC 1084
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1061 TGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCC 1120

Qy       1085 GCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATG 1144
```

```
Db      1121 GCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATG 1180
Qy      1145 GCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACT 1204
Db      1181 GCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACT 1240
Qy      1205 ATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAG 1264
Db      1241 ATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAG 1300
Qy      1265 ATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAG 1324
Db      1301 ATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAG 1360
Qy      1325 GAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTG 1384
Db      1361 GAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTG 1420
Qy      1385 CAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGGTG 1444
Db      1421 CAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGGTG 1480
Qy      1445 CCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTGCA 1504
Db      1481 CCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTGCA 1540
Qy      1505 GTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAAAG 1564
Db      1541 GTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAAAG 1600
Qy      1565 AACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGA 1624
Db      1601 AACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGA 1660
Qy      1625 AATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAA 1684
Db      1661 AATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAA 1720
Qy      1685 AAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGA 1744
Db      1721 AAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGA 1780
Qy      1745 AGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAG 1804
Db      1781 AGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAG 1840
Qy      1805 TAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACC 1864
Db      1841 TAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACC 1900
Qy      1865 TAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCCTGTC 1924
Db      1901 TAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCCTGTC 1960
Qy      1925 CAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCAGC 1984
Db      1961 CAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCAGC 2020
Qy      1985 GGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTCTGTC 2044
Db      2021 GGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTCTGTC 2080
Qy      2045 AGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGTGAACGTTCACCTTAAT 2104
Db      2081 AGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGTGAACGTTCACCTTAAT 2140
Qy      2105 GATTGATTCTTTAATCTCTGTTTTCACTCTCAGGCTCTGGTAAGTATTTGTATTCTCTTC 2164
Db      2141 GATTGATTCTTTAATCTCTGTTGTCACTCTCAGGCTCTGGTAAGTATTCGTATTCTCTTC 2200
Qy      2165 ATCCCAGTCTGATTGCATAGCCACACTGCCCGGCACGCCACATCCACCCCTGTCTGCACA 2224
Db      2201 ATCCCAGTCTGATTGCATAGCCACACTGCCCGGCACGCCACATCCACCCCTGTCTGCACA 2260
Qy      2225 TGAGTTGTTCTGACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGGCCAG 2284
Db      2261 TGAGTTGTTCTGACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGGCCAG 2320
Qy      2285 CACATGAAAGCATCACTTCTTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTGCATC 2344
Db      2321 CACATGAAAGCATCACTTCTTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTGCATC 2380
Qy      2345 TGATTTTCAGGTGTACATTTATTTTTGACTGGGCAGATAGGGGATTTTTTTTTTTTTCCA 2404
Db      2381 TGATTTTCAGGTGTACATTTATTTTTGACTGGGCAGATAGGGGA--TTTTTTTTTTTCCA 2438
Qy      2405 TGTCCGATTCACACGCTACACACCCACATGAACACATTCGAACTTCGAAGG-CACACACT 2463
Db      2439 TGTCCGATTCACACGCTACACACCCACATGAACACATTCGAACTTCGAAGGCCACACACT 2498
Qy      2464 CCTGCTTCATAGGCCCCACGGTAAGTGAGTTCACACCTAGAACACTGTCCTGACCGCAGG 2523
Db      2499 CCTGCTTCATAGGCCCCACGGTAAGTGAGTTCACACCTAGAACACTGTCCTGACCGCAGG 2558
Qy      2524 ACGCGTGCCTTGGACTTGGTATTCTACATGTGACTGGCTTTCTTGCCCTCGTCTCTTGAA 2583
Db      2559 ACGCGTGCCTTGGACTTGGTATTCTACATGTGACTGGCTTTCTTGCCCTCGTCTCTTGAA 2618
Qy      2584 TGTTTAGACTCTTAAGATCATATCCTG 2610
Db      2619 TGTTTAGACTCTTAAGATCATATCCTG 2645
```

RESULT 7
AK002127

LOCUS	AK002127 2772 bp mRNA linear PRI 01-AUG-2002
DEFINITION	Homo sapiens cDNA FLJ11265 fis, clone PLACE1009158.
ACCESSION	AK002127
VERSION	AK002127.1 GI:7023814
KEYWORDS	oligo capping; fis (full insert sequence).
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1
AUTHORS	Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H., Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K., Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
TITLE	NEDO human cDNA sequencing project
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 2772)
AUTHORS	Isogai,T. and Otsuki,T.
TITLE	Direct Submission
JOURNAL	Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

FEATURES Location/Qualifiers

```
     source          1. .2772
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="PLACE1009158"
                     /tissue_type="placenta"
                     /clone_lib="PLACE1"
```

```
                     /note="cloning vector: pME18SFL3"
     CDS             301. .1989
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAA92094.1"
                     /db_xref="GI:7023815"
                     /translation="MDDKGDPSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDAE
                     ADPLEPPPPQQQLGLSLRRSGRQPKRTERVEQFLTIARRRGRRSMPVSLEDSGEPTSC
                     PATDAETASEGSVESASETRSGPQSASTAVKERPASSEKVKGGDDHDDTSDSDSDGLT
                     LKELQNRLRRKREQEPTERPLKGIQSRLRKKRREEGPAETVGSEASDTVEGVLPSKQE
                     PENDQGVVSQAGKDDRESKLEGKAAQDIKDEEPEDLGRPKPECEGYDPNALYCICRQP
                     HNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETHSETAD
                     QQEAKWRPGDADGTDCTSIGTIEQKSSEDQGIKGRIEKAANPSGKKKLKIFQPVIEAP
                     GASKCIGPGCCHVAQPDSVYCSNDCILKHAAATMKFLSSGKEQKPKPKEKMKMKPEKP
                     SLPKCGAQAGIKISSVHKRPAPEKKETTVKKAVVVPARSEALGKEAACESSTPSWASD
                     HNYNAVKPEKTAAPSPSLLYKCMYHLGVGLLDPSRSFWIAIPWACPGLGVAALC"
ORIGIN

  Query Match           96.4%;  Score 2516.2;  DB 9;  Length 2772;
  Best Local Similarity 99.8%;  Pred. No. 0;
  Matches 2541;  Conservative    0;  Mismatches    3;  Indels    3;  Gaps    2;

Qy         65 GGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTG 124
Db        101 GGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTG 160

Qy        125 TTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCC 184
Db        161 TTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCC 220

Qy        185 AGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCAAG 244
Db        221 AGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCAAG 280

Qy        245 CTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTA 304
Db        281 CTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTA 340

Qy        305 AGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTA 364
Db        341 AGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTA 400

Qy        365 TCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCAC 424
Db        401 TCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCAC 460

Qy        425 AGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCG 484
Db        461 AGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCG 520

Qy        485 TGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGG 544
Db        521 TGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGG 580

Qy        545 AGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCA 604
Db        581 AGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCA 640

Qy        605 GCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGG 664
Db        641 GCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGG 700

Qy        665 AACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATA 724
Db        701 AACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATA 760

Qy        725 GTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAAC 784
Db        761 GTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAAC 820

Qy        785 AGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGG 844
Db        821 AGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGG 880

Qy        845 AGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGC 904
Db        881 AGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGC 940

Qy        905 CCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACA 964
Db        941 CCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACA 1000

Qy        965 GAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACT 1024
Db       1001 GAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGAAGACT 1060

Qy       1025 TGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCC 1084
Db       1061 TGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCC 1120

Qy       1085 GCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATG 1144
Db       1121 GCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATG 1180

Qy       1145 GCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACT 1204
Db       1181 GCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACT 1240

Qy       1205 ATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAG 1264
Db       1241 ATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAG 1300

Qy       1265 ATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAG 1324
Db       1301 ATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAG 1360

Qy       1325 GAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTG 1384
Db       1361 GAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTG 1420

Qy       1385 CAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGGTG 1444
Db       1421 CAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGGTG 1480

Qy       1445 CCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTGCA 1504
Db       1481 CCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTGCA 1540

Qy       1505 GTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAAAG 1564
Db       1541 GTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAAAG 1600

Qy       1565 AACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGA 1624
Db       1601 AACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGA 1660

Qy       1625 AATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAA 1684
Db       1661 AATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAA 1720

Qy       1685 AAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGA 1744
Db       1721 AAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGA 1780

Qy       1745 AGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAG 1804
Db       1781 AGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAG 1840

Qy       1805 TAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACC 1864
Db       1841 TAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACC 1900

Qy       1865 TAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCCTGGGCCTGTC 1924
Db       1901 TAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCCTGGGCCTGTC 1960
```

```
Qy       1925 CAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCAGC 1984
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1961 CAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCAGC 2020

Qy       1985 GGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTCTGTC 2044
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2021 GGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTCTGTC 2080

Qy       2045 AGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGTGAACGTTCACCTTAAT 2104
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2081 AGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGTGAACGTTCACCTTAAT 2140

Qy       2105 GATTGATTCTTTAATCTCTGTTTTCACTCTCAGGCTCTGGTAAGTATTTGTATTCTCTTC 2164
             ||||||||||||||||||||||| |||||||||||||||||||||||| ||||||||||||
Db       2141 GATTGATTCTTTAATCTCTGTTGTCACTCTCAGGCTCTGGTAAGTATTCGTATTCTCTTC 2200

Qy       2165 ATCCCAGTCTGATTGCATAGCCACACTGCCCGGCACGCCACATCCACCCCTGTCTGCACA 2224
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2201 ATCCCAGTCTGATTGCATAGCCACACTGCCCGGCACGCCACATCCACCCCTGTCTGCACA 2260

Qy       2225 TGAGTTGTTCTGACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGGCCAG 2284
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2261 TGAGTTGTTCTGACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGGCCAG 2320

Qy       2285 CACATGAAAGCATCACTTCTTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTGCATC 2344
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2321 CACATGAAAGCATCACTTCTTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTGCATC 2380

Qy       2345 TGATTTTCAGGTGTACATTTATTTTTGACTGGGCAGATAGGGGATTTTTTTTTTTTTCCA 2404
             |||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Db       2381 TGATTTTCAGGTGTACATTTATTTTTGACTGGGCAGATAGGGGA--TTTTTTTTTTTCCA 2438

Qy       2405 TGTCCGATTCACACGCTACACACCCACATGAACACATTCGAACTTCGAAGG-CACACACT 2463
             |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db       2439 TGTCCGATTCACACGCTACACACCCACATGAACACATTCGAACTTCGAAGGCCACACACT 2498

Qy       2464 CCTGCTTCATAGGCCCCACGGTAAGTGAGTTCACACCTAGAACACTGTCCTGACCGCAGG 2523
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2499 CCTGCTTCATAGGCCCCACGGTAAGTGAGTTCACACCTAGAACACTGTCCTGACCGCAGG 2558

Qy       2524 ACGCGTGCCTTGGACTTGGTATTCTACATGTGACTGGCTTTCTTGCCCTCGTCTCTTGAA 2583
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2559 ACGCGTGCCTTGGACTTGGTATTCTACATGTGACTGGCTTTCTTGCCCTCGTCTCTTGAA 2618

Qy       2584 TGTTTAGACTCTTAAGATCATATCCTG 2610
             ||||||||||||||||||||||||||||
Db       2619 TGTTTAGACTCTTAAGATCATATCCTG 2645
```

```
RESULT 8
BC000770
LOCUS       BC000770               2407 bp    mRNA    linear   PRI 12-NOV-2003
DEFINITION  Homo sapiens death associated transcription factor 1, transcript
            variant 3, mRNA (cDNA clone MGC:3257 IMAGE:3506207), complete cds.
ACCESSION   BC000770
VERSION     BC000770.2  GI:33990873
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2407)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2407)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2000) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     On Aug 20, 2003 this sequence version replaced gi:12653952.
            Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 7 Row: i Column: 2
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, Similarity but not identity to protein.
FEATURES             Location/Qualifiers
     source          1. .2407
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:3257 IMAGE:3506207"
                     /tissue_type="Kidney, renal cell adenocarcinoma"
                     /clone_lib="NIH_MGC_14"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     gene            1. .2407
                     /gene="DATF1"
                     /note="synonyms: DIO1, FLJ11265, KIAA0333"
                     /db_xref="LocusID:11083"
                     /db_xref="MIM:604140"
     CDS             201. .1835
                     /codon_start=1
                     /product="death associated transcription factor 1, isoform
                     b"
                     /protein_id="AAH00770.1"
                     /db_xref="GI:12653953"
                     /db_xref="LocusID:11083"
                     /translation="MDDKGDPSNEEAPKAIKPTSKEPRKTWGPRRTTIAKREGAGDAE
                     ADPLEPPPPQQQLGLSLRRSGRQPKRTERVEQFLTIARRRGRRSMPVSLEDSGEPTSC
                     PATDAETASEGSVESASETRSGPQSASTAVKERPASSEKVKGGDDHDDTSDSDSDGLT
                     LKELQNRLRRKREQEPTERPLKGIQSRLRKKRREEGPAETVGSEASDTVEGVLPSKQE
                     PENDQGVVSQAGKDDRESKLEGKAAQDIKDEEPGDLGRPKPECEGYDPNALYCICRQP
                     HNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETHSETAD
                     QQEAKWRPGDADGTDCTSIGTIEQKSSEDQGIKGRIEKAANPSGKKKLKIFQPVIEAP
                     GASKCIGPGCCHVAQPDSVYCSNDCILKHAAATMKFLSSGKEQKPKPKEKMKMKPEKP
                     SLPKCGAQAGIKISSVHKRPAPEKKETTVKKAVVVPARSEALGKEAACESSTPSWASD
                     HNYNAVKPEKTAAPSPSLLYKCSGKYLYSLHPSLIA"
     misc_feature    1011. .1160
                     /note="PHD; Region: PHD-finger. PHD folds into an
                     interleaved type of Zn-finger chelating 2 Zn ions in a
```

similar manner to that of the RING and FYVE domains"
/db_xref="CDD:pfam00628"
ORIGIN

Query Match 74.4%; Score 1941; DB 9; Length 2407;
Best Local Similarity 88.7%; Pred. No. 0;
Matches 2259; Conservative 0; Mismatches 0; Indels 288; Gaps 3;

```
Qy        65 GGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTG 124
Db         1 GGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTG 60

Qy       125 TTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCC 184
Db        61 TTACTCGTGAACAGTGGCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCC 120

Qy       185 AGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCAAG 244
Db       121 AGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTTCAGGGAAAAGTGTCCAAG 180

Qy       245 CTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTA 304
Db       181 CTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTA 240

Qy       305 AGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTA 364
Db       241 AGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTA 300

Qy       365 TCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCAC 424
Db       301 TCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCAC 360

Qy       425 AGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCG 484
Db       361 AGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCG 420

Qy       485 TGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGG 544
Db       421 TGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGG 480

Qy       545 AGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCA 604
Db       481 AGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCA 540

Qy       605 GCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGG 664
Db       541 GCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGG 600

Qy       665 AACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATA 724
Db       601 AACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATA 660

Qy       725 GTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAAC 784
Db       661 GTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAAC 720

Qy       785 AGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGG 844
Db       721 AGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGG 780

Qy       845 AGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGC 904
Db       781 AGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGC 840

Qy       905 CCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACA 964
Db       841 CCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACA 900

Qy       965 GAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACT 1024
Db       901 GAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACT 960

Qy      1025 TGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCC 1084
```

```
Db       961 TGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCC 1020

Qy      1085 GCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATG 1144
Db       1021 GCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATG 1080

Qy      1145 GCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACT 1204
Db       1081 GCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACT 1140

Qy      1205 ATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAG 1264
Db       1141 ATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAG 1200

Qy      1265 ATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAG 1324
Db       1201 ATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAG 1260

Qy      1325 GAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTG 1384
Db       1261 GAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTG 1320

Qy      1385 CAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGGTG 1444
Db       1321 CAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGGTG 1380

Qy      1445 CCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTGCA 1504
Db       1381 CCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTGCA 1440

Qy      1505 GTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAAAG 1564
Db       1441 GTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAAAG 1500

Qy      1565 AACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGA 1624
Db       1501 AACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGA 1560

Qy      1625 AATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAA 1684
Db       1561 AATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAA 1620

Qy      1685 AAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGA 1744
Db       1621 AAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGA 1680

Qy      1745 AGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAG 1804
Db       1681 AGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAG 1740

Qy      1805 TAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACC 1864
Db       1741 TAAAGCCAGAAAAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAAT------------ 1788

Qy      1865 TAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCCTGTC 1924
Db       1789 ------------------------------------------------------------ 1788

Qy      1925 CAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCAGC 1984
Db       1789 ------------------------------------------------------------ 1788

Qy      1985 GGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTCTGTC 2044
Db       1789 ------------------------------------------------------------ 1788

Qy      2045 AGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGTGAACGTTCACCTTAAT 2104
Db       1789 ------------------------------------------------------------ 1788

Qy      2105 GATTGATTCTTTAATCTCTGTTTTCACTCTCAGGCTCTGGTAAGTATTTGTATTCTCTTC 2164
```

```
Db      1789 ---------------------------------GCTCTGGTAAGTATTTGTATTCTCTTC 1815

Qy      2165 ATCCCAGTCTGATTGCATAGCCACACTGCCCGGCACGCCACATCCACCCCTGTCTGCACA 2224
Db      1816 ATCCCAGTCTGATTGCATAGCCACACTGCCCGGCACGCCACATCCACCCCTGTCTGCACA 1875

Qy      2225 TGAGTTGTTCTGACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGGCCAG 2284
Db      1876 TGAGTTGTTCTGACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGGCCAG 1935

Qy      2285 CACATGAAAGCATCACTTCTTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTGCATC 2344
Db      1936 CACATGAAAGCATCACTTCTTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTGCATC 1995

Qy      2345 TGATTTTCAGGTGTACATTTATTTTTGACTGGGCAGATAGGGGATTTTTTTTTTTTTCCA 2404
Db      1996 TGATTTTCAGGTGTACATTTATTTTTGACTGGGCAGATAGGGGA--TTTTTTTTTTTCCA 2053

Qy      2405 TGTCCGATTCACACGCTACACACCCACATGAACACATTCGAACTTCGAAGG-CACACACT 2463
Db      2054 TGTCCGATTCACACGCTACACACCCACATGAACACATTCGAACTTCGAAGGCCACACACT 2113

Qy      2464 CCTGCTTCATAGGCCCCACGGTAAGTGAGTTCACACCTAGAACACTGTCCTGACCGCAGG 2523
Db      2114 CCTGCTTCATAGGCCCCACGGTAAGTGAGTTCACACCTAGAACACTGTCCTGACCGCAGG 2173

Qy      2524 ACGCGTGCCTTGGACTTGGTATTCTACATGTGACTGGCTTTCTTGCCCTCGTCTCTTGAA 2583
Db      2174 ACGCGTGCCTTGGACTTGGTATTCTACATGTGACTGGCTTTCTTGCCCTCGTCTCTTGAA 2233

Qy      2584 TGTTTAGACTCTTAAGATCATATCCTG 2610
Db      2234 TGTTTAGACTCTTAAGATCATATCCTG 2260
```

```
RESULT 9
MMU238332
LOCUS       MMU238332             2602 bp    mRNA    linear   ROD 09-JUL-1999
DEFINITION  Mus musculus mRNA for death inducer-obliterator-1 (Dio-1).
ACCESSION   AJ238332
VERSION     AJ238332.1  GI:5457402
KEYWORDS    death inducer-obliterator-1; Dio-1 gene.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Garcia-Domingo,D., Leonardo,E., Grandien,A., Martinez,P.,
            Albar,J.P., Izpisua-Belmonte,J.C. and Martinez-A,C.
  TITLE     DIO-1 is a gene involved in onset of apoptosis in vitro, whose
            misexpression disrupts limb development
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (14), 7992-7997 (1999)
  MEDLINE   99324176
   PUBMED   10393935
REFERENCE   2  (bases 1 to 2602)
  AUTHORS   Garcia-Domingo,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-1999) Garcia-Domingo D., Department of Immunology
            and Oncology, Centro Nacional de Biotecnologia/CSIC, Campus
            Cantoblanco UAM, Ctra. de Colmenar Viejo Km 16, Madrid, 28049,
            SPAIN
FEATURES             Location/Qualifiers
     source          1..2602
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /cell_line="WOL-1"
                     /cell_type="pre-BI"
     gene            1..2602
                     /gene="Dio-1"
     CDS             232..>2073
                     /gene="Dio-1"
                     /function="apoptosis inducer"
                     /codon_start=1
                     /evidence=experimental
                     /product="death inducer-obliterator-1"
                     /protein_id="CAB48401.1"
                     /db_xref="GI:5457403"
                     /db_xref="GOA:Q8C9B9"
                     /db_xref="GOA:Q9WV00"
                     /db_xref="SWISS-PROT:Q8C9B9"
                     /translation="MDDKGHLSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTE
                     ADPSEQPQQHNLSLRRSGRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTV
                     TDVETASEGSVESSSEIRSGPVSDSLGKEHPASSEKAKGGEEEEDTSDSDSDGLTLKE
                     LQNRLRRKREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSAEQDRPLCKQEPEA
                     SQGPVSQSETDDIENQLEGKATQGNTEENPREAGKPKPECEVYDPNALYCICRQPHNN
                     RFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQD
                     SGCRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLKIFQPVVEAPGAP
                     KCIGPGCSSVAQPDSVYCSNDCILKHAAATMRFLSSGKEQKTKPKEKVKTKPEKFSLP
                     KCSVQVGIKISSVHKRLASEKRENPVKKVMLASRSETSGKEAACESSTPSWASDHNYN
                     AVKPEKPEKPTALSPTLLSKCTYHPKAGFPGPSHHLGGCLGLSRTRVLGVLVLIVASS
                     SLPARSRYQDASGPQVFLPSLWSLSGWFLKSCVGLMLEAISYFSFRPW"
ORIGIN

  Query Match             39.2%;  Score 1023.6;  DB 10;  Length 2602;
  Best Local Similarity   71.9%;  Pred. No. 8.6e-264;
  Matches 1494;  Conservative    0;  Mismatches  534;  Indels   50;  Gaps   10;
```

```
Qy         21 TCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGCCTGCGGAGCTTACTCCACGG 80
Db          1 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 60

Qy         81 GAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTGTTACTCGTGAACAGTG 140
Db         61 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 120

Qy        141 GCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCCAGAGGTTTCGTCTGCC 200
Db        121 ACTGACAACAGTGGGGTGAG--GCTTGGCCGTCTGCTTGACC-----------TGGCCC 166

Qy        201 AGGGTTTTTGGTTGTATTTAGGATTT-CAGGGAAAAGTGTCCAAGCTTTCAGTGTTGGAG 259
Db        167 CAGGTCTATAATTTTATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGA 226

Qy        260 CAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTAAGGCCATCAAACCCA 319
Db        227 CAGGTATGGATGATAAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCA 286

Qy        320 CCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTATCGCCAAGCGAGAGG 379
Db        287 CCAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGG 346

Qy        380 GCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCACAGCAGCAGCTGGGCC 439
Db        347 GTGCAGGAGACACGGAGGCGGACCCCAGTGAGCAGCAACC------ACAGCAGCATAACC 400

Qy        440 TGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGA 499
Db        401 TCTCCCTGCGCCGCAGTGGACGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTA 460

Qy        500 CCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGC 559
Db        461 CCACGGTTCGGCGCCGAGGGAAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGC 520

Qy        560 CCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCAGCGTGGAAAGCGCTT 619
Db        521 CCACATCTTCCACAGTCACTGATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTT 580

Qy        620 CTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGGAACGACCAGCCTCTT 679
Db        581 CTGAGATCAGAAGTGGCCCTGTATCTGACTCCTTA---GGGAAAGAACATCCTGCCTCTT 637

Qy        680 CTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGGCC 739
Db        638 CTGAAAAGGCAAAAGGAGGTGAAGAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCC 697
```

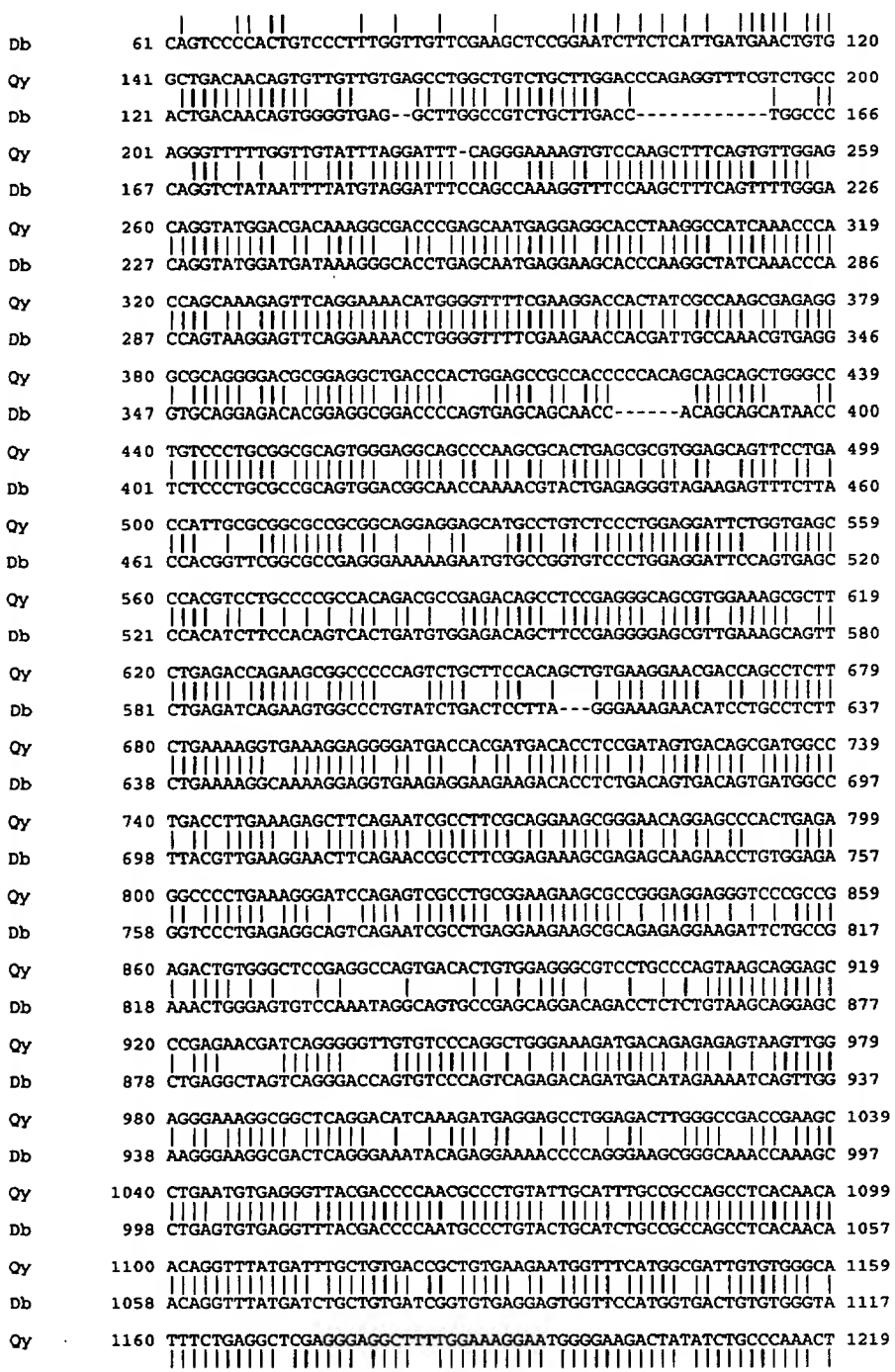
```
Qy       740 TGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGA 799
Db       698 TTACGTTGAAGGAACTTCAGAACCGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGA 757

Qy       800 GGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCG 859
Db       758 GGTCCCTGAGAGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCG 817

Qy       860 AGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGC 919
Db       818 AAACTGGGAGTGTCCAAATAGGCAGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGC 877

Qy       920 CCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGG 979
Db       878 CTGAGGCTAGTCAGGGACCAGTGTCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGG 937

Qy       980 AGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACTTGGGCCGACCGAAGC 1039
Db       938 AAGGGAAGGCGACTCAGGGAAATACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGC 997

Qy      1040 CTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACA 1099
Db       998 CTGAGTGTGAGGTTTACGACCCCAATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACA 1057

Qy      1100 ACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCA 1159
Db      1058 ACAGGTTTATGATCTGCTGTGATCGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTA 1117

Qy      1160 TTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACT 1219
Db      1118 TTTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATT 1177

Qy      1220 GCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTA 1279
Db      1178 GCACCATTTTGCAAGTGCAGGATGAGACAAACGGTAGCGCCACCAATGAGCAGGACTCTG 1237

Qy      1280 AATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAGGAACAATAGAGCAGA 1339
Db      1238 GGTGCAGATCTGTGGGTGCTGATGGCACAGACTGCACAAGCATAGGGACAGTAGAGCAGA 1297

Qy      1340 AGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTGCAAATCCAAGTGGCA 1399
Db      1298 AGTCCGGAGAAGACCAGGGCATAAAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCA 1357

Qy      1400 AGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGGTGCCTCAAAATGTATTG 1459
Db      1358 AGAAAAAACTCAAGATATTCCAGCCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTG 1417

Qy      1460 GCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTGCAGTAATGACTGTATCC 1519
Db      1418 GCCCTGGGTGTTCCAGTGTAGCACAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTC 1477

Qy      1520 TCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGC 1579
Db      1478 TCAAACACGCAGCAGCTACCATGAGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAAC 1537

Qy      1580 CTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGG 1639
Db      1538 CCAAGGAAAAGGTCAAGACGAAGCCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGG 1597

Qy      1640 CAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAAAAAAAAGAGACCACAG 1699
Db      1598 TGGGGATTAAAATCTCTTCTGTGCACAAGAGACTAGCGTCAGAGAAAAGGGAAAACCCAG 1657

Qy      1700 TGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTG 1759
Db      1658 TGAAGAA---AGTGATGCTGGCTTCCAGGAGTGAGACTTCTGGGAAGGAGGCAGCCTGTG 1714

Qy      1760 AGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAGTAAAGCCAGAAAAGA 1819
Db      1715 AGAGCAGCACACCATCCTGGGCAAGTGACCACAACTACAATGCTGTGAAGCCAGAGAAGC 1774

Qy      1820 CTG---------CTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACCTAGGGG 1870
```

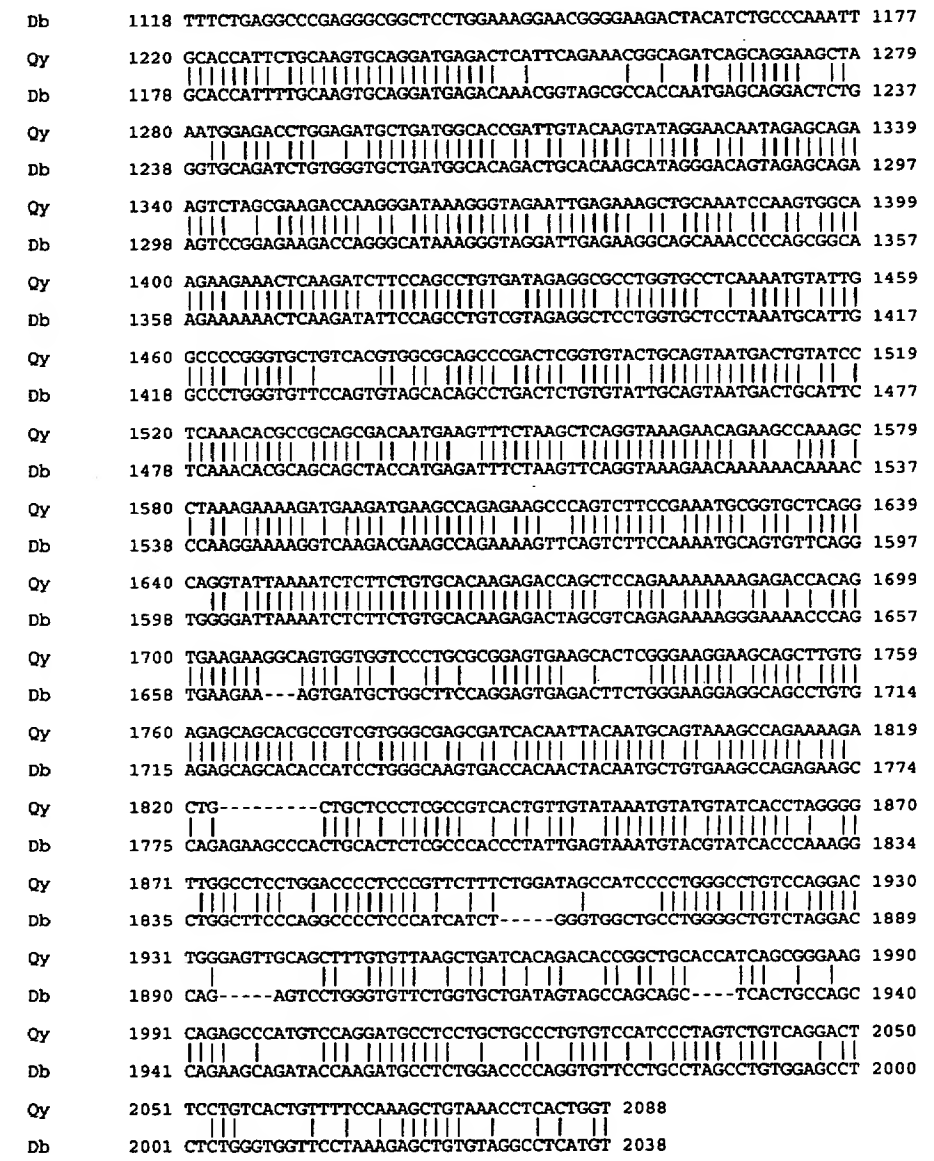
```
Db      1775 CAGAGAAGCCCACTGCACTCTCGCCCACCCTATTGAGTAAATGTACGTATCACCCAAAGG 1834

Qy      1871 TTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCCTGTCCAGGAC 1930
Db      1835 CTGGCTTCCCAGGCCCCTCCCATCATCT-----GGGTGGCTGCCTGGGGCTGTCTAGGAC 1889

Qy      1931 TGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCAGCGGGAAG 1990
Db      1890 CAG-----AGTCCTGGGTGTTCTGGTGCTGATAGTAGCCAGCAGC----TCACTGCCAGC 1940

Qy      1991 CAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTCTGTCAGGACT 2050
Db      1941 CAGAAGCAGATACCAAGATGCCTCTGGACCCCAGGTGTTCCTGCCTAGCCTGTGGAGCCT 2000

Qy      2051 TCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGT 2088
Db      2001 CTCTGGGTGGTTCCTAAAGAGCTGTGTAGGCCTCATGT 2038
```

RESULT 10
BD244713
LOCUS BD244713 2867 bp DNA linear PAT 17-JUL-2003
DEFINITION Genes encoding for the human and murine death inducer-obliterator-1.
ACCESSION BD244713
VERSION BD244713.1 GI:33054483
KEYWORDS JP 2002526040-A/2.
SOURCE Mus sp.
ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2867)
AUTHORS Alonso,C.M., Domingo,D.G., Grandien,A., Leonardo,E. and Martinez,P.
TITLE Genes encoding for the human and murine death inducer-obliterator-1
JOURNAL Patent: JP 2002526040-A 2 20-AUG-2002;
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS
COMMENT OS Mus sp. (murine)
PN JP 2002526040-A/2
PD 20-AUG-2002
PF 10-SEP-1999 JP 2000570314
PR 10-SEP-1998 SE 9803069-5,17-SEP-1998 US 60/100873 PI
CARLOS MARTINEZ ALONSO,DAVID GARCIA DOMINGO,ALP GRANDIEN, PI
ESTHER LEONARDO,
PI PEDRO MARTINEZ
PC C12N15/09,A61K38/00,A61K45/00,A61K48/00,A61P3/10,A61P17/00, PC
A61P29/00,
PC A61P35/00,A61P37/06,A61P43/00,C07K14/47,C07K16/18,C12N1/21, PC
C12N5/10,
PC C12P21/02,C12Q1/02,C12Q1/68,G01N33/15,G01N33/50//C12P21/08, PC
C12N15/00,
PC C12N5/00,A61K37/02
CC Genes encoding for the human and murine death inducer- CC
obliterator-1
FH Key Location/Qualifiers
FT source 1. .2867
FT /organism='Mus sp. (murine)'.
FEATURES Location/Qualifiers
source 1. .2867
/organism="Mus sp."
/mol_type="genomic DNA"
/db_xref="taxon:10095"
ORIGIN

Query Match 39.2%; Score 1023.6; DB 6; Length 2867;
Best Local Similarity 71.9%; Pred. No. 8.7e-264;
Matches 1494; Conservative 0; Mismatches 534; Indels 50; Gaps 10;

```
Qy        21 TCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGCCTGCGGAGCTTACTCCACGG 80
Db         1 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 60

Qy        81 GAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTGTTACTCGTGAACAGTG 140
```





```
RESULT 11
AX023370
LOCUS       AX023370               2867 bp    DNA     linear   PAT 15-SEP-2000
DEFINITION  Sequence 2 from Patent WO0015787.
ACCESSION   AX023370
VERSION     AX023370.1  GI:10183782
KEYWORDS    .
SOURCE      Mus sp.
  ORGANISM  Mus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Leonardo,E., Martinez,P., Alonso,C.M., Domingo,D.G. and Grandien,A.
  TITLE     Genes encoding for the human and murine death in
            ducer-obliterator-1
  JOURNAL   Patent: WO 0015787-A 2 23-MAR-2000;
            LEONARDO ESTHER (ES) ; MARTINEZ PEDRO (ES) ; CONSEJO SUPERIOR
            INVESTIGACION (ES) ; DOMINGO DAVID GARCIA (ES) ; MARTINEZ ALONSO
            CARLOS (ES) ; BANNERMAN DAVID GARDNER (GB) ; GRANDIEN ALF (SE)
FEATURES             Location/Qualifiers
     source          1. .2867
                     /organism="Mus sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:10095"
ORIGIN

  Query Match           39.2%;  Score 1023.6;  DB 6;  Length 2867;
  Best Local Similarity  71.9%;  Pred. No. 8.7e-264;
  Matches 1494;  Conservative    0;  Mismatches  534;  Indels   50;  Gaps   10;

Qy         21 TCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGCCTGCGGAGCTTACTCCACGG 80
Db          1 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 60

Qy         81 GAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTGTTACTCGTGAACAGTG 140
Db         61 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 120

Qy        141 GCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCCAGAGGTTTCGTCTGCC 200
Db        121 ACTGACAACAGTGGGGTGAG--GCTTGGCCGTCTGCTTGACC-----------TGGCCC 166

Qy        201 AGGGTTTTTGGTTGTATTTAGGATTT-CAGGGAAAAGTGTCCAAGCTTTCAGTGTTGGAG 259
Db        167 CAGGTCTATAATTTTATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGA 226

Qy        260 CAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTAAGGCCATCAAACCCA 319
Db        227 CAGGTATGGATGATAAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCA 286

Qy        320 CCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTATCGCCAAGCGAGAGG 379
Db        287 CCAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGG 346

Qy        380 GCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCACAGCAGCAGCTGGGCC 439
Db        347 GTGCAGGAGACACGGAGGCGGACCCCAGTGAGCAGCAACC------ACAGCAGCATAACC 400

Qy        440 TGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGA 499
Db        401 TCTCCCTGCGCCGCAGTGGACGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTA 460

Qy        500 CCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGC 559
Db        461 CCACGGTTCGGCGCCGAGGGAAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGC 520

Qy        560 CCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCAGCGTGGAAAGCGCTT 619
Db        521 CCACATCTTCCACAGTCACTGATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTT 580

Qy        620 CTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGGAACGACCAGCCTCTT 679
Db        581 CTGAGATCAGAAGTGGCCCTGTATCTGACTCCTTA---GGGAAAGAACATCCTGCCTCTT 637

Qy        680 CTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGGCC 739
Db        638 CTGAAAAGGCAAAAGGAGGTGAAGAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCC 697

Qy        740 TGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGA 799
Db        698 TTACGTTGAAGGAACTTCAGAACCGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGA 757

Qy        800 GGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCG 859
```

```
Db        758 GGTCCCTGAGAGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCG 817

Qy        860 AGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGC 919
Db        818 AAACTGGGAGTGTCCAAATAGGCAGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGC 877

Qy        920 CCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGG 979
Db        878 CTGAGGCTAGTCAGGGACCAGTGTCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGG 937

Qy        980 AGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACTTGGGCCGACCGAAGC 1039
Db        938 AAGGGAAGGCGACTCAGGGAAATACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGC 997

Qy       1040 CTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACA 1099
Db        998 CTGAGTGTGAGGTTTACGACCCCAATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACA 1057

Qy       1100 ACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCA 1159
Db       1058 ACAGGTTTATGATCTGCTGTGATCGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTA 1117

Qy       1160 TTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACT 1219
Db       1118 TTTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATT 1177

Qy       1220 GCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTA 1279
Db       1178 GCACCATTTTGCAAGTGCAGGATGAGACAAACGGTAGCGCCACCAATGAGCAGGACTCTG 1237

Qy       1280 AATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAGGAACAATAGAGCAGA 1339
Db       1238 GGTGCAGATCTGTGGGTGCTGATGGCACAGACTGCACAAGCATAGGGACAGTAGAGCAGA 1297

Qy       1340 AGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTGCAAATCCAAGTGGCA 1399
Db       1298 AGTCCGGAGAAGACCAGGGCATAAAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCA 1357

Qy       1400 AGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGGTGCCTCAAAATGTATTG 1459
Db       1358 AGAAAAAACTCAAGATATTCCAGCCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTG 1417

Qy       1460 GCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTGCAGTAATGACTGTATCC 1519
Db       1418 GCCCTGGGTGTTCCAGTGTAGCACAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTC 1477

Qy       1520 TCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGC 1579
Db       1478 TCAAACACGCAGCAGCTACCATGAGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAAC 1537

Qy       1580 CTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGG 1639
Db       1538 CCAAGGAAAAGGTCAAGACGAAGCCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGG 1597

Qy       1640 CAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAAAAAAAGAGACCACAG 1699
Db       1598 TGGGGATTAAAATCTCTTCTGTGCACAAGAGACTAGCGTCAGAGAAAAGGGAAAACCCAG 1657

Qy       1700 TGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTG 1759
Db       1658 TGAAGAA---AGTGATGCTGGCTTCCAGGAGTGAGACTTCTGGGAAGGAGGCAGCCTGTG 1714

Qy       1760 AGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAGTAAAGCCAGAAAAGA 1819
Db       1715 AGAGCAGCACACCATCCTGGGCAAGTGACCACAACTACAATGCTGTGAAGCCAGAGAAGC 1774

Qy       1820 CTG---------CTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACCTAGGGG 1870
Db       1775 CAGAGAAGCCCACTGCACTCTCGCCCACCCTATTGAGTAAATGTACGTATCACCCAAAGG 1834

Qy       1871 TTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCCTGTCCAGGAC 1930
```

```
Db       1835 CTGGCTTCCCAGGCCCCTCCCATCATCT-----GGGTGGCTGCCTGGGGCTGTCTAGGAC 1889

Qy       1931 TGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCAGCGGGAAG 1990

Db       1890 CAG-----AGTCCTGGGTGTTCTGGTGCTGATAGTAGCCAGCAGC----TCACTGCCAGC 1940

Qy       1991 CAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTCTGTCAGGACT 2050

Db       1941 CAGAAGCAGATACCAAGATGCCTCTGGACCCCAGGTGTTCCTGCCTAGCCTGTGGAGCCT 2000

Qy       2051 TCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGT 2088

Db       2001 CTCTGGGTGGTTCCTAAAGAGCTGTGTAGGCCTCATGT 2038
```

```
RESULT 12
AK129117
LOCUS       AK129117                4906 bp    mRNA    linear   ROD 21-NOV-2003
DEFINITION  Mus musculus mRNA for mKIAA0333 protein.
ACCESSION   AK129117
VERSION     AK129117.1  GI:37359897
KEYWORDS    FLI_CDNA.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Okazaki,N., Kikuno,R., Ohara,R., Inamoto,S., Koseki,H., Hiraoka,S.,
            Saga,Y., Nagase,T., Ohara,O. and Koga,H.
  TITLE     Prediction of the coding sequences of mouse homologues of KIAA
            gene: III. the complete nucleotide sequences of 500 mouse
            KIAA-homologous cDNAs identified by screening of terminal sequences
            of cDNA clones randomly sampled from size-fractionated libraries
  JOURNAL   DNA Res. 10 (4), 167-180 (2003)
  MEDLINE   22977043
   PUBMED   14621295
REFERENCE   2  (bases 1 to 4906)
  AUTHORS   Okazaki,N., Kikuno,R., Nagase,T., Ohara,O. and Koga,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-2003) Hisashi Koga, Kazusa DNA Research
            Institute, Laboratory for Genome Informatics; 2-6-7
            Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan
            (E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)
COMMENT     The CREATE program supported by Japan science and technology
            corporation; cDNA full insert sequencing: Kazusa DNA Research
            Institute; cDNA library construction, clone selection and 5'- &
            3'-end one pass sequencing.
FEATURES             Location/Qualifiers
     source          1. .4906
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="mbg16015"
                     /tissue_type="brain"
                     /dev_stage="adult"
                     /note="vector:modified pBC SK+"
     gene            1. .4906
                     /gene="mKIAA0333"
     CDS             <224. .3829
                     /gene="mKIAA0333"
                     /note="CDS is predicted by in silico analysis. Start codon
                     is not identified."
                     /codon_start=1
                     /evidence=not_experimental
                     /product="mKIAA0333 protein"
                     /protein_id="BAC97927.1"
                     /db_xref="GI:37359898"
                     /translation="FYVGFPAKGFQAFSFGTGMDDKGHLSNEEAPKAIKPTSKEFRKT
                     WGFRRTTIAKREGAGDTEVDPSEQQPQQHNLSLRRSGRQPKRTERVEEPLTTVRRRGK
                     KNVPVSLEDSSEPTSSTVTDVETASEGSVESSSEIRSGPVSDSLGKEHPASSEKAKGG
                     EEEEDTSDSDSDGLTLKELQNRLRRKREQEPVERSLRGSQNRLRKKRREEDSAETGSV
                     QIGSAEQDRPLCKQEPEASQGPVSQSETDDIENQLEGKATQGNTEENPREAGKPKPEC
                     EVYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNC
```

```
                     TILQVQDETNGSATDEQDSGCRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPS
                     GKKKLKIFQPVVEAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAAATMRFLSSGKEQ
                     KTKPKEKVKTKPEKFSLPKCSVQVGIKISSVHKRLASEKRENPVKKVMLASRSETSGK
                     EAACESSTPSWASDHNYNAVKPEKPEKPTALSPTLLSKSMKDDRRVEDRTMAAVTIPK
                     KALPSASLVGRQTSPRNLVPKKLPPYSNMAGAKPAIKKLPSGFKGTIPKRPWPSATLS
                     GTSARQAGPTPMTAASKKLPGSAAVVGVTRKPMSANVPAASPAPGRLGPVSPAPSQPN
                     SQIRQNIRRSLKEILWKRVNDSDDLIMTENEVGKIALHIEKEMFNLFQVTDNRYKSKY
                     RSIMFNLKDPKNQGLFHRVLREEISLAKLVRMKPEELVSKELSMWTEKPTKSVIESRT
                     KLLNESKKNTTKPETIPDMEDSPPVSDSEEQQESVRAAPEKSAAPLLDVFSSMLKDTT
                     SQHRAHLFDLNCKICTGQVPSSEDEPAPKKQKLSASSKKEDFKPRHDSSPPNAVPNTA
                     DEGIADTLPENASEPDPESTSSLNQERKCFPESPGDSHPEPSSLGGLSPSSASGGSGV
                     VTTVTMSGRDPRTALSGSCTVTASMAAHLDNSQASETKLDMIKPALTSAVVPKSILAK
                     PSSSPDPRYLSVPPSPSISESRSPPEGDTTLFLSRLNTIWKGFINMQSVAKFVTKAYP
                     VSGCLDYLSEDLPDTIHIGGRIAPKTVWDYVGKLKSSVSKELCLIRFHPATEEEEVAY
                     ISLYSYFSSRGRFGVVANNNRHVKDLYLIPLSAKDPVPSKLLPFEGPGKHPVSGR"
ORIGIN

  Query Match           37.7%;  Score 985;  DB 10;  Length 4906;
  Best Local Similarity 76.0%;  Pred. No. 2.3e-253;
  Matches 1288;  Conservative    0;  Mismatches  385;  Indels   22;  Gaps    5;

Qy       175 GCTTGGACCCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTT-CAGGGAA 233

Db       187 GCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTTTATGTAGGATTTCCAGCCAA 246

Qy       234 AAGTGTCCAAGCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGA 293

Db       247 AGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGATAAAGGGCACCTGAGCAATGA 306

Qy       294 GGAGGCACCTAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCG 353

Db       307 GGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCG 366

Qy       354 AAGGACCACTATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCC 413

Db       367 AAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACGGAGGTGGACCCCAGTGAGCA 426

Qy       414 GCCACCCCCACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCG 473

Db       427 GCAACC------ACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGACGGCAACCAAAACG 480

Qy       474 CACTGAGCGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCC 533

Db       481 TACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGGAAAAAGAATGTGCC 540

Qy       534 TGTCTCCCTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGC 593

Db       541 GGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACTGATGTGGAGACAGC 600

Qy       594 CTCCGAGGGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCAC 653

Db       601 TTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCTGTATCTGACTCCTT 660

Qy       654 AGCTGTGAAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGA 713

Db       661 A---GGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAAGAGGAAGAAGA 717

Qy       714 CACCTCCGATAGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAG 773

Db       718 CACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAACCGCCTTCGGAG 777

Qy       774 GAAGCGGGAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAA 833

Db       778 AAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAATCGCCTGAGGAA 837

Qy       834 GAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGA 893

Db       838 GAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGCAGTGCCGAGCA 897

Qy       894 GGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGG 953

Db       898 GGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTGTCCCAGTCAGA 957
```

```
Qy        954 GAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGA 1013
              || |||||||| ||| | | ||||||| || |||||| ||||||  |  | ||| ||  |
Db        958 GACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAATACAGAGGAAAA 1017

Qy       1014 GCCTGGAGACTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTA 1073
              ||  | ||    |||| ||| |||||||| ||||||| |||||||||||| |||||||||
Db       1018 CCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCCAATGCCCTGTA 1077

Qy       1074 TTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGA 1133
              ||||| |||||||||||||||||||||||||||||||||| |||||||| || ||||| ||
Db       1078 CTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGATCGGTGTGAGGA 1137

Qy       1134 ATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAA 1193
              ||||| ||||| || |||||||| |||||||||| |||||| ||||  ||||||||||
Db       1138 GTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAA 1197

Qy       1194 TGGGGAAGACTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTC 1253
              |||||||||||| |||||||||| ||||||||| ||||||||||||||||||||  |
Db       1198 CGGGGAAGACTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGATGAGACAAACGG 1257

Qy       1254 AGAAACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTG 1313
                 |  ||| |||||||  ||    || ||| |||  | ||||||||||||| || ||
Db       1258 TAGCGCCACCGATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGATGGCACAGACTG 1317

Qy       1314 TACAAGTATAGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAAT 1373
              ||||| ||||| ||| ||||||||||||||  | |||||||| || |||||||||| ||
Db       1318 CACAAGCATAGGGACAGTAGAGCAGAAGTCCGGAGAAGACCAGGGCATAAAGGGTAGGAT 1377

Qy       1374 TGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGA 1433
              |||||| || ||||| || || |||||||| |||||||||| |||||||||| ||||
Db       1378 TGAGAAGGCAGCAAACCCCAGCGGCAAGAAAAAACTCAAGATATTCCAGCCTGTCGTAGA 1437

Qy       1434 GGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTC 1493
              ||| |||||||| | ||||| |||||||| ||||| |      || || ||||| |||||
Db       1438 GGCTCCTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCACAGCCTGACTC 1497

Qy       1494 GGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAG 1553
              ||||| |||||||||||||| || |||||||||| ||||| || ||||  ||||||||
Db       1498 TGTGTATTGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATGAGATTTCTAAG 1557

Qy       1554 CTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCC 1613
              |||||||||||||| ||  |||| || || |||||| | |||| |||||||||| |||
Db       1558 TTCAGGTAAAGAACAAAAAACAAAACCCAAGGAAAAGGTCAAGACGAAGCCAGAAAAGTT 1617

Qy       1614 CAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACC 1673
              ||||||||| |||||| ||| ||||||  || |||||||||||||||||||||||||||
Db       1618 CAGTCTTCCAAAATGCAGTGTTCAGGTGGGGATTAAAATCTCTTCTGTGCACAAGAGACT 1677

Qy       1674 AGCTCCAGAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGA 1733
              |||  |||| ||||  || | | |||||||||||   |||| || |  || |  |||||||
Db       1678 AGCGTCAGAGAAAAGGGAAAACCCAGTGAAGAA---AGTGATGCTGGCTTCCAGGAGTGA 1734

Qy       1734 AGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAA 1793
                |     |||||||| ||||| |||||||||||||| || || ||||| || || |||||
Db       1735 GACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGCACACCATCCTGGGCAAGTGACCACAA 1794

Qy       1794 TTACAATGCAGTAAAGCCAGAAAAGACTG---------CTGCTCCCTCGCCGTCACTGTT 1844
              |||||||| || |||||||| ||| | |          |||| | ||||||  | || ||
Db       1795 CTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAGCCCACTGCACTCTCGCCCACCCTATT 1854

Qy       1845 GTATAAATGTATGTA 1859
              |  |||||  ||| |
Db       1855 GAGTAAATCCATGAA 1869
```

RESULT 13
HS885L7/c

```
LOCUS       HS885L7              160241 bp    DNA     linear   PRI 19-APR-2001
DEFINITION  Human DNA sequence from clone RP5-885L7 on chromosome 20q13.2-13.33
            Contains ESTs, STSs, GSSs and eight CpG islands. Contains the 3'
            end of the NTSR1 gene for high affinity neurotensin receptor 1, a
            putative novel gene, a novel gene similar to a fly gene, the gene
            for opioid growth factor receptor (7-60 protein), the COL9A3 gene
            for collagen IX alpha 3, a putative novel gene similar to a fly
            gene, the TCFL5 gene for basic helix-loop-helix transcription
            factor-like 5, an ARF4 (ADP-ribosylation factor 4) pseudogene, a
            novel gene and the 3' end of the DATF1 gene encoding death
            associated transcription factor 1, complete sequence.
ACCESSION   AL035669
VERSION     AL035669.43  GI:8979786
KEYWORDS    HTG; 7-60; ARF4; COL9A3; collagen; CpG island; DIO-1; KIAA0333;
            neurotensin; NTSR1; opioid growth factor receptor; TCFL5.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 160241)
  AUTHORS   Smith,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 8, 2000 this sequence version replaced gi:8919619.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            This sequence is the entire insert of clone RP5-885L7 The true left
            end of clone RP4-563E14 is at 140192 in this sequence. The true
            right end of clone RP11-93B14 is at 44512 in this sequence. This
            sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. RP5-885L7 is from the
            library RPCI-5 constructed by the group of Pieter de Jong. For
            further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .160241
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /map="q13.2-13.33"
                     /clone="RP5-885L7"
                     /clone_lib="RPCI-5"
     repeat_region   2. .64
                     /note="L1MA9 repeat: matches 6210. .6254 of consensus"
     repeat_region   65. .353
                     /note="AluJo repeat: matches 1. .287 of consensus"
     repeat_region   354. .397
                     /note="L1MA9 repeat: matches 6254. .6302 of consensus"
     gene            3862. .11948
                     /gene="NTSR1"
     mRNA            join(<3862. .4063,7443. .7533,9195. .11948)
                     /gene="NTSR1"
                     /product="dJ885L7.1.1 (neurotensin receptor 1 (high
                     affinity), isoform 1)"
                     /note="match: cDNAs: Em:AB017027 Em:E11480 Em:X70070
                     match: ESTs: Em:AA078742 Em:U47726 Em:H29546 Em:AA157908
                     Em:AA074404 Em:H29638 Em:AA079834 Em:AA157852"
```

```
                /evidence=not_experimental
CDS             join(<3862. .4063,7443. .7533,9195. .9444)
                /gene="NTSR1"
                /note="Continues in bA93B14 (Em:AL357033) as bA93B14.3
                match: proteins: Tr:O88319 Sw:P30989 Sw:P20789"
                /codon_start=1
                /evidence=not_experimental
                /product="dJ885L7.1.1 (neurotensin receptor 1 (high
                affinity), isoform 1)"
                /protein_id="CAC12747.1"
                /db_xref="GI:12313997"
                /db_xref="GOA:P30989"
                /db_xref="SWISS-PROT:P30989"
                /translation="VNTFMSFIFPMVVISVLNTIIANKLTVMVRQAAEQGQVCTVGGE
                HSTFSMAIEPGRVQALRHGVRVLRAVVIAFVVCWLPYHVRRLMFCYISDEQWTPFLYD
                FYHYFYMVTNALFYVSSTINPILYNLVSANFRHIFLATLACLCPVWRRRRKRPAPSRK
                ADSVSSNHTLSSNATRETLY"
repeat_region   4198. .5157
                /note="12 copies 80 mer 61% conserved"
misc_feature    complement(4223. .5153)
                /note="Tandem repeat. region contains two forced joins and
                single clone. Assembly consistent with restriction
                digest."
repeat_region   4229. .5114
                /note="443 copies 2 mer cc 59% conserved"
repeat_region   4231. .5166
                /note="234 copies 4 mer tccc 60% conserved"
repeat_region   4240. .5151
                /note="19 copies 48 mer 61% conserved"
repeat_region   4248. .5159
                /note="38 copies 24 mer 61% conserved"
repeat_region   4264. .5163
                /note="15 copies 60 mer 61% conserved"
repeat_region   4286. .5152
                /note="17 copies 51 mer 59% conserved"
repeat_region   4308. .5171
                /note="27 copies 32 mer 61% conserved"
repeat_region   4317. .5156
                /note="21 copies 40 mer 62% conserved"
repeat_region   4344. .5142
                /note="17 copies 47 mer 59% conserved"
repeat_region   4348. .5145
                /note="42 copies 19 mer 60% conserved"
repeat_region   4431. .5144
                /note="21 copies 34 mer 60% conserved"
repeat_region   4456. .5148
                /note="21 copies 33 mer 60% conserved"
repeat_region   4486. .5151
                /note="9 copies 74 mer 69% conserved"
repeat_region   4530. .5087
                /note="18 copies 31 mer 71% conserved"
repeat_region   4533. .4822
                /note="5 copies 58 mer 77% conserved"
repeat_region   4720. .5137
                /note="11 copies 38 mer 61% conserved"
repeat_region   4833. .5064
                /note="4 copies 58 mer 79% conserved"
repeat_region   5051. .5164
                /note="19 copies 6 mer tctctc 63% conserved"
mRNA            join(5808. .5912,7443. .7533,9195. .11948)
                /gene="NTSR1"
                /product="dJ885L7.1.2 (neurotensin receptor 1 (high
                affinity), isoform 2)"
                /note="match: ESTs: Em:AL537553"
                /evidence=not_experimental
repeat_region   5934. .6864
                /note="L1M4 repeat: matches 4509. .5488 of consensus"
repeat_region   6974. .7049
                /note="4 copies 19 mer 80% conserved"
repeat_region   7224. .7451
                /note="6 copies 38 mer 98% conserved"
repeat_region   7892. .8074
                /note="MER68A repeat: matches 68. .210 of consensus"
repeat_region   8674. .8966
                /note="AluY repeat: matches 5. .296 of consensus"
repeat_region   10876. .10962
                /note="MIR repeat: matches 86. .162 of consensus"
polyA_signal    11918. .11923
                /gene="NTSR1"
polyA_site      11948
                /gene="NTSR1"
repeat_region   11974. .12025
                /note="13 copies 4 mer ctca 73% conserved"
repeat_region   14394. .14478
                /note="5 copies 17 mer 77% conserved"
misc_feature    14766. .14966
                /note="match: GSS: Em:B99679"
repeat_region   15824. .15891
                /note="17 copies 4 mer tccc 67% conserved"
repeat_region   16266. .16577
                /note="AluSq repeat: matches 1. .313 of consensus"
repeat_region   16578. .16866
                /note="17 copies 17 mer 65% conserved"
repeat_region   16581. .16844
                /note="11 copies 24 mer 68% conserved"
repeat_region   16582. .16891
                /note="10 copies 31 mer 66% conserved"
repeat_region   16583. .16879
                /note="9 copies 33 mer 66% conserved"
repeat_region   16585. .16840
                /note="16 copies 16 mer 66% conserved"
repeat_region   16586. .16875
                /note="5 copies 58 mer 72% conserved"
repeat_region   16590. .16874
                /note="15 copies 19 mer 64% conserved"
repeat_region   16593. .16872
                /note="7 copies 40 mer 67% conserved"
repeat_region   16598. .16837
                /note="5 copies 48 mer 73% conserved"
repeat_region   16603. .16842
                /note="3 copies 80 mer 77% conserved"
repeat_region   16638. .16901
                /note="44 copies 6 mer aaagaa 62% conserved"
repeat_region   16664. .16885
                /note="3 copies 74 mer 78% conserved"
repeat_region   16670. .16904
                /note="5 copies 47 mer 77% conserved"
repeat_region   16848. .16899
                /note="13 copies 4 mer gaag 100% conserved"
repeat_region   17287. .17411
                /note="MER5A repeat: matches 35. .165 of consensus"

Query Match           36.3%;  Score 948.4;  DB 9;  Length 160241;
Best Local Similarity   99.6%;  Pred. No. 3e-243;
Matches  972;  Conservative    0;  Mismatches    1;  Indels    3;  Gaps    2;

Qy       1636 CAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAAAAAAAAGAGACC 1695
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     155280 CAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAAAAAAAAGAGACC 155221

Qy       1696 ACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCT 1755
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     155220 ACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCT 155161

Qy       1756 TGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAGTAAAGCCAGAA 1815
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     155160 TGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAGTAAAGCCAGAA 155101

Qy       1816 AAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACCTAGGGGTTGGC 1875
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     155100 AAGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACCTAGGGGTTGGC 155041

Qy       1876 CTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCCTGTCCAGGACTGGGA 1935
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     155040 CTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCCTGTCCAGGACTGGGA 154981
```

```
Qy       1936 GTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCAGCGGGAAGCAGAG 1995
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     154980 GTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCAGCGGGAAGCAGAG 154921

Qy       1996 CCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTCTGTCAGGACTTCCTG 2055
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     154920 CCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTCTGTCAGGACTTCCTG 154861

Qy       2056 TCACTGTTTTCCAAAGCTGTAAACCTCACTGGTGAACGTTCACCTTAATGATTGATTCTT 2115
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     154860 TCACTGTTTTCCAAAGCTGTAAACCTCACTGGTGAACGTTCACCTTAATGATTGATTCTT 154801

Qy       2116 TAATCTCTGTTTTCACTCTCAGGCTCTGGTAAGTATTTGTATTCTCTTCATCCCAGTCTG 2175
             |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     154800 TAATCTCTGTTTTCACTCTCAGGCTCTGGTAAGTATTCGTATTCTCTTCATCCCAGTCTG 154741

Qy       2176 ATTGCATAGCCACACTGCCCGGCACGCCACATCCACCCCTGTCTGCACATGAGTTGTTCT 2235
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     154740 ATTGCATAGCCACACTGCCCGGCACGCCACATCCACCCCTGTCTGCACATGAGTTGTTCT 154681

Qy       2236 GACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGGCCAGCACATGAAAGC 2295
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     154680 GACAACAGCGCTGTATACGCTTCAGTTTTTCCACATTGTCCACGGCCAGCACATGAAAGC 154621

Qy       2296 ATCACTTCTTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTGCATCTGATTTTCAGG 2355
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     154620 ATCACTTCTTTTTTATGTTGTGGGAATCTTTGCAAGTTAGTGTTGCATCTGATTTTCAGG 154561

Qy       2356 TGTACATTTATTTTTGACTGGGCAGATAGGGGATTTTTTTTTTTTTCCATGTCCGATTCA 2415
             |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db     154560 TGTACATTTATTTTTGACTGGGCAGATAGGGGA--TTTTTTTTTTTCCATGTCCGATTCA 154503

Qy       2416 CACGCTACACACCCACATGAACACATTCGAACTTCGAAGG-CACACACTCCTGCTTCATA 2474
             |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     154502 CACGCTACACACCCACATGAACACATTCGAACTTCGAAGGCCACACACTCCTGCTTCATA 154443

Qy       2475 GGCCCCACGGTAAGTGAGTTCACACCTAGAACACTGTCCTGACCGCAGGACGCGTGCCTT 2534
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     154442 GGCCCCACGGTAAGTGAGTTCACACCTAGAACACTGTCCTGACCGCAGGACGCGTGCCTT 154383

Qy       2535 GGACTTGGTATTCTACATGTGACTGGCTTTCTTGCCCTCGTCTCTTGAATGTTTAGACTC 2594
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     154382 GGACTTGGTATTCTACATGTGACTGGCTTTCTTGCCCTCGTCTCTTGAATGTTTAGACTC 154323

Qy       2595 TTAAGATCATATCCTG 2610
             ||||||||||||||||
Db     154322 TTAAGATCATATCCTG 154307
```

RESULT 14
AB002331
LOCUS AB002331 6692 bp mRNA linear PRI 06-OCT-2001
DEFINITION Human mRNA for KIAA0333 gene, partial cds.
ACCESSION AB002331
VERSION AB002331.1 GI:2224606
KEYWORDS KIAA0333.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Nagase,T., Ishikawa,K., Nakajima,D., Ohira,M., Seki,N., Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro
JOURNAL DNA Res. 4 (2), 141-150 (1997)
MEDLINE 97349984
PUBMED 9205841
REFERENCE 2 (bases 1 to 6692)
AUTHORS Ohara,O., Nagase,T., Kikuno,R. and Nomura,N.
TITLE Direct Submission
JOURNAL Submitted (28-MAR-1997) Osamu Ohara, Kazusa DNA Research Institute; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913)

```
FEATURES             Location/Qualifiers
     source          1. .6692
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HG0988"
                     /sex="male"
                     /tissue_type="brain"
                     /clone_lib="pBluescriptII SK plus"
     gene            1. .6692
                     /gene="KIAA0333"
     CDS             <1. .2978
                     /gene="KIAA0333"
                     /codon_start=3
                     /protein_id="BAA20791.1"
                     /db_xref="GI:2224607"
                     /translation="ESKLEGKAAQDIKDEEPGDLGRPKPECEGYDPNALYCICRQPHN
                     NRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETHSETADQQ
                     EAKWRPGDADGTDCTSIGTIEQKSSEDQGIKGRIEKAANPSGKKKLKIFQPGPGPVPT
                     QLPVLWQVLEIAVSRSISAFTLLHCISCKVIEAPGASKCIGPGCCHVAQPDSVYCSND
                     CILKHAAATMKFLSSGKEQKPKPKEKMKMKPEKPSLPKCGAQAGIKISSVHKRPAPEK
                     KETTVKKAVVVPARSEALGKEAACESSTPSWASDHNYNAVKPEKTAAPSPSLLYKSTK
                     EDRRSEEKAAATAASKKTAPPGSTVGKQPAPRNLVPKKSSFANVAAATPAIKKPPSGF
                     KGTIPKRPWLSATPSSGASAARQAGPAPAAATAASKKFPGSAALVGAVRKPVVPSVPM
                     ASPAPGRLGAMSAAPSQPNSQIRQNIRRSLKEILWKRVNDSDDLIMTENEVGKIALHI
                     EKEMFNLFQVTDNRYKSKYRSIMFNLKDPKNQGLFHRVLREEISLAKLVRLKPEELVS
                     KELSTWKERPARSVMESRTKLHNESKKTAPRQEAIPDLEDSPPVSDSEEQQESARAVP
                     EKSTAPLLDVFSSMLKDTTSQHRAHLFDLNCKICTGQVPSAEDEPAPKKQKLSASVKK
                     EDLKSKHDSSAPDPAPDSADEVMPEAVPEVASEPGLESASHPNVDRTYPPGPPGDGHP
                     EPSPLEDLSPCPASCGSGVVTTVTVSGRDPRTAPSSSCTAVASAASRPDSTHMVEARQ
                     DVPKPVLTSVMVPKSILAKPSSSPDPRYLSVPPSPNISTSESRSPPEGDTTLFLSRLS
                     TIWKGFINMQSVAKPVTKAYPVSGCFDYLSEDLPDTIHIGGRIAPKTVWDYVGKLKSS
                     VSKELCLIRFHPATEEEEVAYISLYSYFSSRGRFGVVANNNRHVKDLYLIPLSAQDPV
                     PSKLLPFEGPGKRRLSGWR"
ORIGIN
```

Query Match 29.5%; Score 770.6; DB 9; Length 6692;
Best Local Similarity 88.8%; Pred. No. 1.3e-195;
Matches 891; Conservative 0; Mismatches 4; Indels 108; Gaps 1;

```
Qy        965 GAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACT 1024
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACT 60

Qy       1025 TGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCC 1084
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCC 120

Qy       1085 GCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATG 1144
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATG 180

Qy       1145 GCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACT 1204
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACT 240

Qy       1205 ATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAG 1264
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAG 300

Qy       1265 ATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAG 1324
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAG 360

Qy       1325 GAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTG 1384
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTG 420

Qy       1385 CAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCT---------- 1425
              |||||||||||||||||||||||||||||||||||||||||||
Db        421 CAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGGTCCCGGGCCTGTCCCCA 480
```

```
Qy       1426 ------------------------------------------------------------ 1425

Db        481 CCCAGCTGCCTGTCCTCTGGCAGGTATTGGAAATTGCTGTGTCTAGAAGCATCTCTGCCT 540

Qy       1426 ----------------------------GTGATAGAGGCGCCTGGTGCCTCAAAATGTA 1456
                                            |||||||||||||||||||||||||||||||
Db        541 TCACACTCTTGCACTGCATTAGCTGCAAGGTGATAGAGGCGCCTGGTGCCTCAAAATGTA 600

Qy       1457 TTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTGCAGTAATGACTGTA 1516
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTGCAGTAATGACTGTA 660

Qy       1517 TCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAAGCCAA 1576
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAAGCCAA 720

Qy       1577 AGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGAAATGCGGTGCTC 1636
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGAAATGCGGTGCTC 780

Qy       1637 AGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAAAAAAAAGAGACCA 1696
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAAAAAAAAGAGACCA 840

Qy       1697 CAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCTT 1756
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCTT 900

Qy       1757 GTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAGTAAAGCCAGAAA 1816
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAGTAAAGCCAGAAA 960

Qy       1817 AGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTA 1859
              |||||||||||||||||||||||||||||||||  | | |
Db        961 AGACTGCTGCTCCCTCGCCGTCACTGTTGTATAAATCCACGAA 1003
```

RESULT 15
HSJ563E14/c

```
LOCUS       HSJ563E14               28519 bp    DNA     linear   PRI 19-APR-2001
DEFINITION  Human DNA sequence from clone RP4-563E14 on chromosome 20 Contains
            the 5' of the DATF1 gene encoding the death associated
            transcription factor 1, the 5' end of a novel gene, ESTs, STSs,
            GSSs and four CpG islands, complete sequence.
ACCESSION   AL117379
VERSION     AL117379.14  GI:7329921
KEYWORDS    HTG; CpG island; DATF1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 28519)
  AUTHORS   Collier,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 26, 2000 this sequence version replaced gi:7327990.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            IMPORTANT: This sequence is not the entire insert of clone
            RP4-563E14 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP11-305P22 is at 28420 in this
            sequence. The true right end of clone RP5-885L7 is at 100 in this
            sequence. This sequence was finished as follows unless otherwise
            noted: all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
            RP4-563E14 is from the library RPCI-4 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2.
FEATURES            Location/Qualifiers
     source         1. .28519
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="20"
                    /clone="RP4-563E14"
                    /clone_lib="RPCI-4"
     gene           complement(1. .26931)
                    /gene="DATF1"
     mRNA           complement(join(<1. .650,3246. .3442,26832. .26931))
                    /gene="DATF1"
                    /product="dJ563E14.1 (Death associated transcription
                    factor 1 (contains KIAA0333) isoform 1)"
                    /note="match: cDNAs: Em:AK002127 Em:AJ238332
                    match: ESTs: Em:AW410673 Em:AA090179 Em:AA343704
                    Em:N55493"
                    /evidence=not_experimental
     CDS            complement(<1. .648)
                    /gene="DATF1"
                    /note="continues in clone dJ885L7 (Em:AL035669) as gene
                    dJ885L7.9
                    match: proteins: Tr:Q9WV00"
                    /codon_start=1
                    /evidence=not_experimental
                    /product="dJ563E14.1 (Death associated transcription
                    factor 1 (contains KIAA0333) isoform 1)"
                    /protein_id="CAB90275.1"
                    /db_xref="GI:7711029"
                    /db_xref="GOA:Q9BTC0"
                    /db_xref="SWISS-PROT:Q9BTC0"
                    /translation="MDDKGDPSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDAE
                    ADPLEPPPPQQQLGLSLRRSGRQPKRTERVEQFLTIARRRGRRSMPVSLEDSGEPTSC
                    PATDAETASEGSVESASETRSGPQSASTAVKERPASSEKVKGGDDHDDTSDSDSDGLT
                    LKELQNRLRRKREQEPTERPLKGIQSRLRKKRREEGPAETVGSEASDTVEGVLPSK"
     misc_feature   39. .561
                    /note="CpG island"
                    /evidence=not_experimental
     mRNA           complement(join(<426. .650,3246. .3442,15475. .>15507))
                    /gene="DATF1"
                    /product="dJ563E14.1 (Death associated transcription
                    factor 1 (contains KIAA0333) isoform 1)"
                    /note="Death associated transcription factor 1
                    (translation of cDNA PLACE1009158 (Em:AK002127)) isoform 2
                    alternatively spliced in 5' UTR
                    match: ESTs: Em:AW410673"
                    /evidence=not_experimental
     misc_feature   complement(1177. .1649)
                    /gene="DATF1"
                    /note="match: GSS: Em:AQ127293"
     repeat_region  1327. .1417
                    /note="MLT1F repeat: matches 439. .535 of consensus"
     repeat_region  1462. .1514
                    /note="AluSc repeat: matches 246. .298 of consensus"
     misc_feature   1749. .2041
                    /note="match: GSS: Em:AQ253731"
```

```
repeat_region   1786. .2093
                /note="AluSq repeat: matches 1. .313 of consensus"
repeat_region   2898. .2953
                /note="MER2 repeat: matches 97. .153 of consensus"
misc_feature    complement(2912. .3305)
                /gene="DATF1"
                /note="match: STS: Em:N55493"
repeat_region   3030. .3097
                /note="MER2 repeat: matches 269. .336 of consensus"
repeat_region   4231. .4366
                /note="AluJb repeat: matches 3. .145 of consensus"
repeat_region   4520. .4828
                /note="AluJo repeat: matches 1. .302 of consensus"
repeat_region   5459. .6077
                /note="Charlie1 repeat: matches 2193. .2759 of consensus"
repeat_region   6078. .6366
                /note="AluJb repeat: matches 1. .266 of consensus"
repeat_region   6367. .6994
                /note="Charlie1 repeat: matches 4. .2193 of consensus"
repeat_region   7348. .7470
                /note="AluJb repeat: matches 136. .259 of consensus"
repeat_region   8302. .8594
                /note="AluSq repeat: matches 1. .301 of consensus"
repeat_region   9647. .9932
                /note="Charlie1b repeat: matches 2. .239 of consensus"
repeat_region   9933. .10231
                /note="AluSq repeat: matches 1. .300 of consensus"
repeat_region   10232. .10392
                /note="Charlie1b repeat: matches 239. .523 of consensus"
repeat_region   10669. .10836
                /note="AluJb repeat: matches 135. .301 of consensus"
misc_feature    10781. .10901
                /note="match: STS: Em:G32962"
misc_feature    10898. .11020
                /note="match: GSS: Em:AQ387275"
repeat_region   10956. .11135
                /note="3 copies 60 mer 78% conserved"
repeat_region   11138. .11305
                /note="14 copies 12 mer 67% conserved"
repeat_region   11146. .11309
                /note="41 copies 4 mer atac 64% conserved"
misc_feature    complement(11948. .12516)
                /gene="DATF1"
                /note="match: GSS: Em:AQ415373"
repeat_region   12058. .12208
                /note="FRAM repeat: matches 1. .144 of consensus"
repeat_region   12213. .12509
                /note="AluSx repeat: matches 1. .301 of consensus"
misc_feature    12555. .13045
                /note="match: GSS: Em:AQ804060"
misc_feature    13880. .14299
                /note="match: GSS: Em:AQ132474"
repeat_region   14246. .14317
                /note="L2 repeat: matches 2581. .2658 of consensus"
misc_feature    14712. .15898
                /note="CpG island"
                /evidence=not_experimental
misc_feature    complement(15147. .15331)
                /gene="DATF1"
                /note="match: GSS: Em:AP190856"
repeat_region   15283. .15310
                /note="7 copies 4 mer cccg 96% conserved"
repeat_region   15515. .15636
                /note="2 copies 61 mer 86% conserved"
misc_feature    18118. .18638
                /note="CpG island"
                /evidence=not_experimental
misc_feature    complement(18967. .19431)
                /gene="DATF1"
                /note="match: GSS: Em:AQ637346"
repeat_region   18989. .19268
                /note="AluSx repeat: matches 2. .312 of consensus"
misc_feature    complement(18996. .19482)
                /gene="DATF1"
                /note="match: GSS: Em:AQ764765"
repeat_region   19976. .20270
                /note="AluJb repeat: matches 1. .302 of consensus"
misc_feature    complement(20348. .20659)
                /gene="DATF1"
                /note="match: GSS: Em:AQ469836"
repeat_region   21084. .21205
                /note="FLAM_C repeat: matches 6. .129 of consensus"
repeat_region   21248. .21364
                /note="AluJb repeat: matches 2. .119 of consensus"
unsure          21547
repeat_region   21967. .22272
                /note="AluSp repeat: matches 1. .308 of consensus"
repeat_region   22632. .22881
                /note="AluJb repeat: matches 9. .263 of consensus"
repeat_region   23208. .23574
                /note="L2 repeat: matches 1520. .1917 of consensus"
repeat_region   24500. .24660
                /note="L2 repeat: matches 2372. .2542 of consensus"
repeat_region   24758. .25032
```

```
Query Match            25.1%;  Score 654.6;  DB 9;  Length 28519;
Best Local Similarity  98.7%;  Pred. No. 2.7e-164;
Matches  660;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       244 GCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCT 303
             |||| || | |       ||||||||||||||||||||||||||||||||||||||||||||
Db       669 GCTTACACTTTACCGCTAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCT 610

Qy       304 AAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACT 363
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       609 AAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACT 550

Qy       364 ATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCA 423
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       549 ATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCA 490

Qy       424 CAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGC 483
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       489 CAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGC 430

Qy       484 GTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTG 543
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       429 GTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTG 370

Qy       544 GAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGC 603
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       369 GAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGC 310

Qy       604 AGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAG 663
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       309 AGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAG 250

Qy       664 GAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGAT 723
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       249 GAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGAT 190

Qy       724 AGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAA 783
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       189 AGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAA 130

Qy       784 CAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGG 843
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       129 CAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGG 70

Qy       844 GAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTG 903
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        69 GAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTG 10

Qy       904 CCCAGTAAG 912
             |||||||||
Db         9 CCCAGTAAG 1
```

Search completed: April 23, 2004, 14:16:14
Job time : 9971.23 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 23, 2004, 04:38:03 ; Search time 1056.35 Seconds
(without alignments)
11529.921 Million cell updates/sec

Title: US-09-787-016A-2
Perfect score: 2867
Sequence: 1 tccgtggtagctctggaaat.....tttgtttccttttgtgctggg 2867

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2867	100.0	2867	3	AAZ90579	Aaz90579 Murine de
	2	1023.6	35.7	2610	3	AAZ90578	Aaz90578 Human dea
	3	1013	35.3	2772	4	AAH15102	Aah15102 Human cDN
	4	949.8	33.1	2085	5	ABV28029	Abv28029 Human pro
	5	949.8	33.1	2085	5	ABV27935	Abv27935 Human pro
	6	949.8	33.1	2085	5	ABV22095	Abv22095 Human pro
	7	948.2	33.1	2085	5	ABV22192	Abv22192 Human pro
	8	719.2	25.1	7290	4	AAK53352	Aak53352 Human pol
	9	719.2	25.1	7838	4	AAK52368	Aak52368 Human pol
	10	692	24.1	2332	5	AAS86576	Aas86576 DNA encod
	11	558	19.5	1764	3	AAC77933	Aac77933 Human can
	12	382.8	13.4	408	3	AAA43780	Aaa43780 Mouse sec
	13	376	13.1	2986	5	AAS86575	Aas86575 DNA encod
	14	347.8	12.1	503	5	AAS86574	Aas86574 DNA encod
	15	322.4	11.2	324	6	ABN21248	Abn21248 Human ORF
	16	288	10.0	478	8	ACH31915	Ach31915 Human end
c	17	257.6	9.0	481	5	ABV42461	Abv42461 Human pro
	18	230.4	8.0	30626	4	AAK67051	Aak67051 Human imm
	19	209.8	7.3	681	4	AAH08005	Aah08005 Human cDN
c	20	166.8	5.8	7838	4	AAK52368	Aak52368 Human pol
	21	165.4	5.8	362	5	ABV12528	Abv12528 Human pro
	22	156.4	5.5	365	5	ABV42575	Abv42575 Human pro
	23	156.4	5.5	365	5	ABV33664	Abv33664 Human pro
	24	116.2	4.1	256	5	ABV03359	Abv03359 Human pro
	25	110.4	3.9	7174	9	ADC37602	Adc37602 Human nuc
	26	82.2	2.9	6051	4	ABL29341	Abl29341 Drosophil
c	27	80.4	2.8	636	5	ABV03229	Abv03229 Human pro
	28	70.2	2.4	8751	4	ABL29340	Abl29340 Drosophil
	29	65	2.3	65	6	ABN52491	Abn52491 Mouse spl
	30	57.8	2.0	434	5	ABA13151	Aba13151 Human ner
	31	57.8	2.0	915	4	AAS26292	Aas26292 Human cDN
	32	57.8	2.0	915	7	ABX73633	Abx73633 Human nov
	33	57.8	2.0	2320	5	ABX71405	Abx71405 Human tes
	34	57.8	2.0	2387	3	AAF22374	Aaf22374 Human sec
	35	56	2.0	10732	3	AAA10594	Aaa10594 Gene enco
	36	55.4	1.9	5975	6	ABL61775	Abl61775 Colon ade
	37	55.4	1.9	6256	4	AAI58390	Aai58390 Human pol
	38	55.4	1.9	6256	8	ADB48359	Adb48359 Novel hum
	39	55.4	1.9	6584	4	AAI60176	Aai60176 Human pol
	40	55.4	1.9	6947	7	ACF34533	Acf34533 Gene enco
	41	55.4	1.9	6948	5	AAH81806	Aah81806 Human dif
	42	55.4	1.9	6996	4	AAK53025	Aak53025 Human pol
	43	55.4	1.9	7354	4	AAK52041	Aak52041 Human pol
	44	53.8	1.9	1992	4	ABL27413	Abl27413 Drosophil
	45	53	1.8	2000	7	ADA71938	Ada71938 Rice gene

ALIGNMENTS

RESULT 1
AAZ90579
ID AAZ90579 standard; cDNA; 2867 BP.
XX
AC AAZ90579;
XX
DT 19-JUN-2000 (first entry)
XX
DE Murine death inducer-obliterator 1 (DIO-1) polypeptide encoding cDNA.
XX
KW Death inducer-obliterator 1; DIO-1; cell death; cancer; tumour; murine;
KW autoimmune disease; cytostatic; immunosuppressive; antidiabetic;
KW antirheumatic; antiinflammatory; antiproliferative; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 232. .2076
FT /*tag= a
FT /product= "DIO-1"
XX
PN WO200015787-A1.
XX
PD 23-MAR-2000.
XX
PF 10-SEP-1999; 99WO-GB003019.
XX
PR 10-SEP-1998; 98SE-00003069.
PR 17-SEP-1998; 98US-0100873P.
XX
PA (CNSJ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.
PA (BANN/) BANNERMAN D G.
XX
PI Alonso CM, Domingo DG, Grandien A, Leonardo E, Martinez P;
XX
DR WPI; 2000-271426/23.
DR P-PSDB; AAY67580.
XX
PT New DNA encoding human and murine death inducer-obliterator 1
PT polypeptides, useful in the treatment of cancer, autoimmune diseases,
PT diabetes, rheumatoid arthritis, benign tumors, malignant tumors and
PT hyperproliferative skin disorders.
XX
PS Claim 3; Fig 1B; 27pp; English.
XX
CC The invention provides nucleic acids encoding the human and murine death

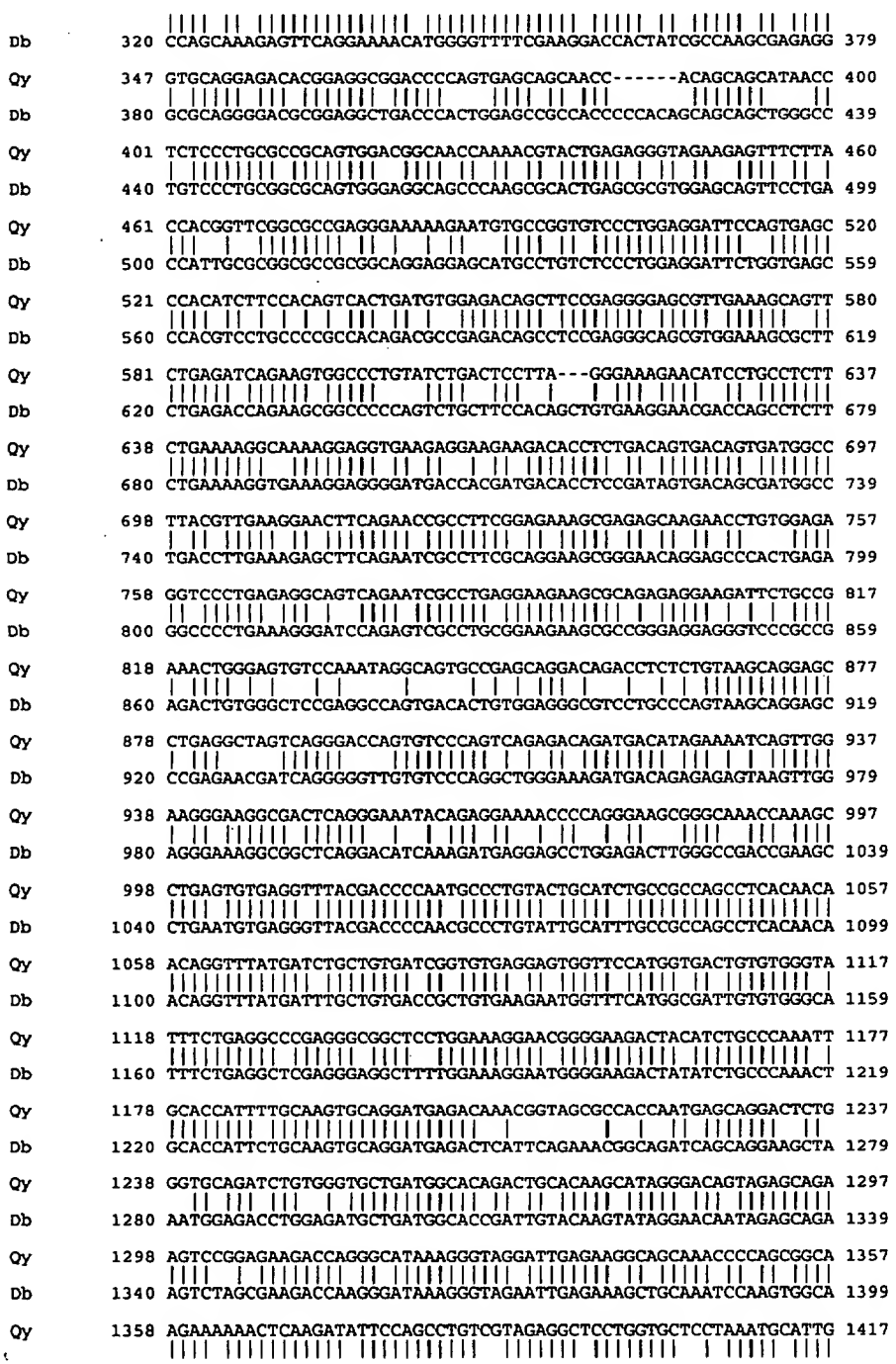
CC inducer-obliterator 1 (DIO-1) polypeptides. The polypeptides can be
CC expressed by standard recombinant methodology. The DIO-1 polypeptides,
CC agonists and antagonists are used as a medicament for treating diseases
CC characterized by an alteration in cell death or by hyperproliferation,
CC e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign
CC tumours, malignant tumours or hyperproliferative skin disorders. They are
CC also useful in the treatment of metabolic, proliferative or inflammatory
CC conditions. The present sequence represents a cDNA encoding the murine
CC DIO-1 polypeptide
XX
SQ Sequence 2867 BP; 763 A; 666 C; 774 G; 664 T; 0 U; 0 Other;

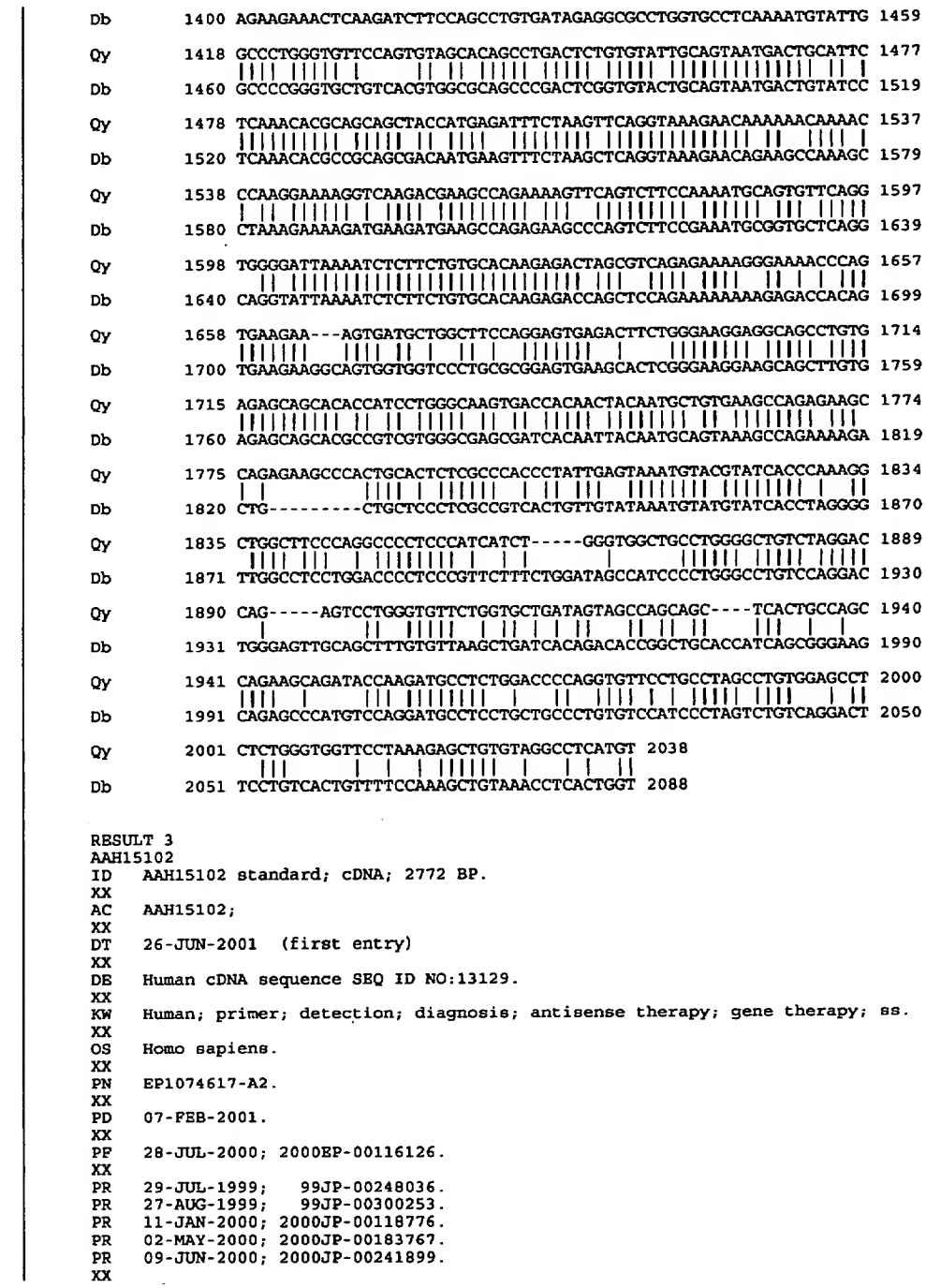
Query Match 100.0%; Score 2867; DB 3; Length 2867;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2867; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 60

Qy         61 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 120

Qy        121 ACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTT 180

Qy        181 TATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGAT 240

Qy        241 AAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTC 300

Qy        301 AGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACG 360

Qy        361 GAGGCGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAGGCGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGA 420

Qy        421 CGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGG 480

Qy        481 AAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACT 540

Qy        541 GATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCT 600

Qy        601 GTATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAA 660

Qy        661 GAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAAC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAAC 720

Qy        721 CGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAAT 780

Qy        781 CGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGC 840

Qy        841 AGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTG 900

Qy        901 TCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAAT 960

Qy        961 ACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCC 1020

Qy       1021 AATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGAT 1080

Qy       1081 CGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTC 1140

Qy       1141 CTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGAT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGAT 1200

Qy       1201 GAGACAAACGGTAGCGCCACCAATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGAT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGACAAACGGTAGCGCCACCAATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGAT 1260

Qy       1261 GGCACAGACTGCACAAGCATAGGGACAGTAGAGCAGAAGTCCGGAGAAGACCAGGGCATA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GGCACAGACTGCACAAGCATAGGGACAGTAGAGCAGAAGTCCGGAGAAGACCAGGGCATA 1320

Qy       1321 AAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCAAGAAAAAACTCAAGATATTCCAG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCAAGAAAAAACTCAAGATATTCCAG 1380

Qy       1381 CCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCA 1440

Qy       1441 CAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATG 1500

Qy       1501 AGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAACCCAAGGAAAAGGTCAAGACGAAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAACCCAAGGAAAAGGTCAAGACGAAG 1560

Qy       1561 CCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTGGGGATTAAAATCTCTTCTGTG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTGGGGATTAAAATCTCTTCTGTG 1620

Qy       1621 CACAAGAGACTAGCGTCAGAGAAAAGGGAAAACCCAGTGAAGAAAGTGATGCTGGCTTCC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CACAAGAGACTAGCGTCAGAGAAAAGGGAAAACCCAGTGAAGAAAGTGATGCTGGCTTCC 1680

Qy       1681 AGGAGTGAGACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGCACACCATCCTGGGCAAGT 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 AGGAGTGAGACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGCACACCATCCTGGGCAAGT 1740

Qy       1741 GACCACAACTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAGCCCACTGCACTCTCGCCC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 GACCACAACTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAGCCCACTGCACTCTCGCCC 1800

Qy       1801 ACCCTATTGAGTAAATGTACGTATCACCCAAAGGCTGGCTTCCCAGGCCCCTCCCATCAT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ACCCTATTGAGTAAATGTACGTATCACCCAAAGGCTGGCTTCCCAGGCCCCTCCCATCAT 1860

Qy       1861 CTGGGTGGCTGCCTGGGGCTGTCTAGGACCAGAGTCCTGGGTGTTCTGGTGCTGATAGTA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 CTGGGTGGCTGCCTGGGGCTGTCTAGGACCAGAGTCCTGGGTGTTCTGGTGCTGATAGTA 1920

Qy       1921 GCCAGCAGCTCACTGCCAGCCAGAAGCAGATACCAAGATGCCTCTGGACCCCAGGTGTTC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1921 GCCAGCAGCTCACTGCCAGCCAGAAGCAGATACCAAGATGCCTCTGGACCCCAGGTGTTC 1980
Qy       1981 CTGCCTAGCCTGTGGAGCCTCTCTGGGTGGTTCCTAAAGAGCTGTGTAGGCCTCATGTTG 2040
Db       1981 CTGCCTAGCCTGTGGAGCCTCTCTGGGTGGTTCCTAAAGAGCTGTGTAGGCCTCATGTTG 2040
Qy       2041 GAGGCAATTTCTTATTTCAGTTTTAGGCCTTGGTAAGTATTTGAACTGCTCTCAACAAGA 2100
Db       2041 GAGGCAATTTCTTATTTCAGTTTTAGGCCTTGGTAAGTATTTGAACTGCTCTCAACAAGA 2100
Qy       2101 TGGGACAACATCAGTGCACTGTTCAGACTCAGTTCAGACTTGAGTTCCTCACAGGACAGC 2160
Db       2101 TGGGACAACATCAGTGCACTGTTCAGACTCAGTTCAGACTTGAGTTCCTCACAGGACAGC 2160
Qy       2161 CCAGCACAAACCATGGAGTTTTCCCACGTTACCTATTTTTCTTCAACATGGTCACCACTG 2220
Db       2161 CCAGCACAAACCATGGAGTTTTCCCACGTTACCTATTTTTCTTCAACATGGTCACCACTG 2220
Qy       2221 CTCTTCTATAGAAAAGTGATTTTTTTTCATGAGTTAGGGCATTTTTGAAAGGACAGGTAGA 2280
Db       2221 CTCTTCTATAGAAAAGTGATTTTTTTTCATGAGTTAGGGCATTTTTGAAAGGACAGGTAGA 2280
Qy       2281 CAATTTGTGTCTATTTCACATACTACACACCTACATAAACAGGTTTGAATTTTGAAGGTC 2340
Db       2281 CAATTTGTGTCTATTTCACATACTACACACCTACATAAACAGGTTTGAATTTTGAAGGTC 2340
Qy       2341 ATCTGCTGCTCAGATCCCATGGTGAGTAATCGTGCTGACCATACACCTACAGCAGTATTC 2400
Db       2341 ATCTGCTGCTCAGATCCCATGGTGAGTAATCGTGCTGACCATACACCTACAGCAGTATTC 2400
Qy       2401 TCATCAAAGTGAGAATTGTGGGTGGTAGCTTCCAGACTCGTAGATGCCCCAGATTTCAAA 2460
Db       2401 TCATCAAAGTGAGAATTGTGGGTGGTAGCTTCCAGACTCGTAGATGCCCCAGATTTCAAA 2460
Qy       2461 TCAGTGTTTATAGTTTGAGTAAACTTTTGAAACCCAAGATTAGATCACTGAGTTTGGAGC 2520
Db       2461 TCAGTGTTTATAGTTTGAGTAAACTTTTGAAACCCAAGATTAGATCACTGAGTTTGGAGC 2520
Qy       2521 CATTCTTGTTTCACTGCTGTACAAGGTACACCTTTCTTACTTTTTGTACAGTGGTTTCAA 2580
Db       2521 CATTCTTGTTTCACTGCTGTACAAGGTACACCTTTCTTACTTTTTGTACAGTGGTTTCAA 2580
Qy       2581 GTGAATTTTCATCTGTGTTCTCATGATAGTGTTTTGTTCCATCTCTGTCCACCCTCAGCC 2640
Db       2581 GTGAATTTTCATCTGTGTTCTCATGATAGTGTTTTGTTCCATCTCTGTCCACCCTCAGCC 2640
Qy       2641 CCAGATAAGGGGGGTCTCATGGCCTAAGCTGGCCTTAAATTTTCTGTAAAGTTGGGGGTCC 2700
Db       2641 CCAGATAAGGGGGTCTCATGGCCTAAGCTGGCCTTAAATTTTCTGTAAAGTTGGGGGTCC 2700
Qy       2701 TCCTGCCCCCACCTGTAAAGAGTGCTGAAACTTACAGGCACACACCTGATGTATGCTGTG 2760
Db       2701 TCCTGCCCCCACCTGTAAAGAGTGCTGAAACTTACAGGCACACACCTGATGTATGCTGTG 2760
Qy       2761 GGGTGTGAACTCAGAGTGTTGGACCAGCAGTCTACCAGCTGAGCTGCAGTTCTAGCCATG 2820
Db       2761 GGGTGTGAACTCAGAGTGTTGGACCAGCAGTCTACCAGCTGAGCTGCAGTTCTAGCCATG 2820
Qy       2821 GTAACTGAAACTGCTGATCTTCTTGTCTTTGTTTCCTTTGTGCTGGG 2867
Db       2821 GTAACTGAAACTGCTGATCTTCTTGTCTTTGTTTCCTTTGTGCTGGG 2867
```

```
RESULT 2
AAZ90578
ID   AAZ90578 standard; cDNA; 2610 BP.
XX
AC   AAZ90578;
XX
DT   19-JUN-2000  (first entry)
XX
DE   Human death inducer-obliterator 1 (DIO-1) polypeptide encoding cDNA.
XX
KW   Death inducer-obliterator 1; DIO-1; cell death; cancer; tumour; human;
KW   autoimmune disease; cytostatic; immunosuppressive; antidiabetic;
KW   antirheumatic; antiinflammatory; antiproliferative; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             265. .1953
FT                   /*tag=  a
FT                   /product= "DIO-1"
XX
PN   WO200015787-A1.
XX
PD   23-MAR-2000.
XX
PF   10-SEP-1999;   99WO-GB003019.
XX
PR   10-SEP-1998;   98SE-00003069.
PR   17-SEP-1998;   98US-0100873P.
XX
PA   (CNSJ ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.
PA   (BANN/) BANNERMAN D G.
XX
PI   Alonso CM,  Domingo DG,  Grandien A,  Leonardo E,  Martinez P;
XX
DR   WPI; 2000-271426/23.
DR   P-PSDB; AAY67579.
XX
PT   New DNA encoding human and murine death inducer-obliterator 1
PT   polypeptides, useful in the treatment of cancer, autoimmune diseases,
PT   diabetes, rheumatoid arthritis, benign tumors, malignant tumors and
PT   hyperproliferative skin disorders.
XX
PS   Claim 1; Fig 1A; 27pp; English.
XX
CC   The invention provides nucleic acids encoding the human and murine death
CC   inducer-obliterator 1 (DIO-1) polypeptides. The polypeptides can be
CC   expressed by standard recombinant methodology. The DIO-1 polypeptides,
CC   agonists and antagonists are used as a medicament for treating diseases
CC   characterized by an alteration in cell death or by hyperproliferation,
CC   e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign
CC   tumours, malignant tumours or hyperproliferative skin disorders. They are
CC   also useful in the treatment of metabolic, proliferative or inflammatory
CC   conditions. The present sequence represents a cDNA encoding the human DIO
CC   -1 polypeptide
XX
SQ   Sequence 2610 BP; 644 A; 667 C; 734 G; 565 T; 0 U; 0 Other;

  Query Match           35.7%;  Score 1023.6;  DB 3;  Length 2610;
  Best Local Similarity  71.9%;  Pred. No. 3.3e-309;
  Matches 1494;  Conservative    0;  Mismatches  534;  Indels   50;  Gaps   10;

Qy         1 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 60
Db        21 TCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGCCTGCGGAGCTTACTCCACGG 80
Qy        61 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 120
Db        81 GAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTGTTACTCGTGAACAGTG 140
Qy       121 ACTGACAACAGTGGGGTGAG--GCTTGGCCGTCTGCTTGACC----------TGGCCC 166
Db       141 GCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCCAGAGGTTTCGTCTGCC 200
Qy       167 CAGGTCTATAATTTTATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGA 226
Db       201 AGGGTTTTTGGTTGTATTTAGGATTT-CAGGGAAAAGTGTCCAAGCTTTCAGTGTTGGAG 259
Qy       227 CAGGTATGGATGATAAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCA 286
Db       260 CAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTAAGGCCATCAAACCCA 319
Qy       287 CCAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGG 346
```

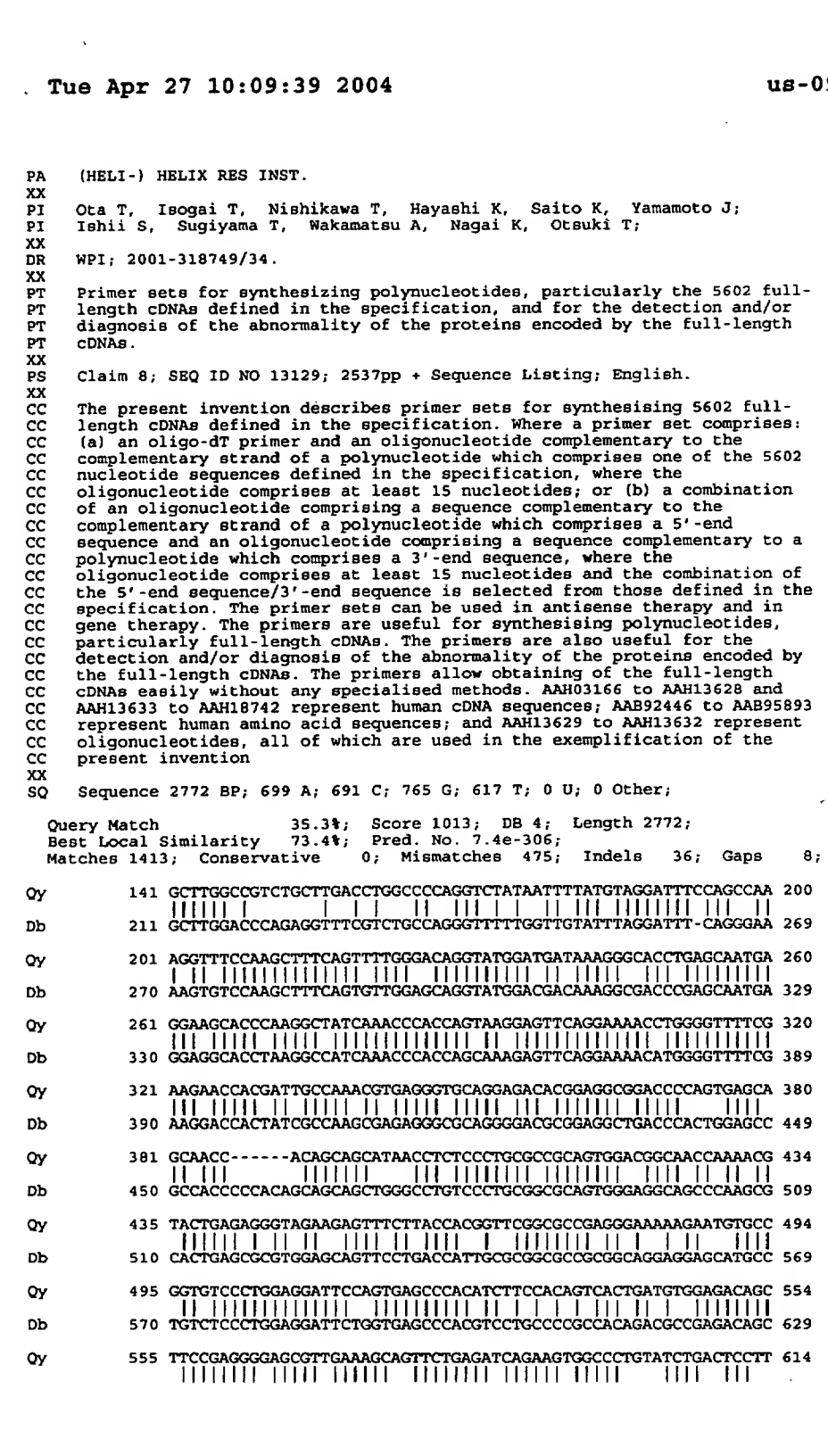




RESULT 3
AAH15102
ID AAH15102 standard; cDNA; 2772 BP.
XX
AC AAH15102;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:13129.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX

PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 8; SEQ ID NO 13129; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesising 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 2772 BP; 699 A; 691 C; 765 G; 617 T; 0 U; 0 Other;

Query Match 35.3%; Score 1013; DB 4; Length 2772;
Best Local Similarity 73.4%; Pred. No. 7.4e-306;
Matches 1413; Conservative 0; Mismatches 475; Indels 36; Gaps 8;

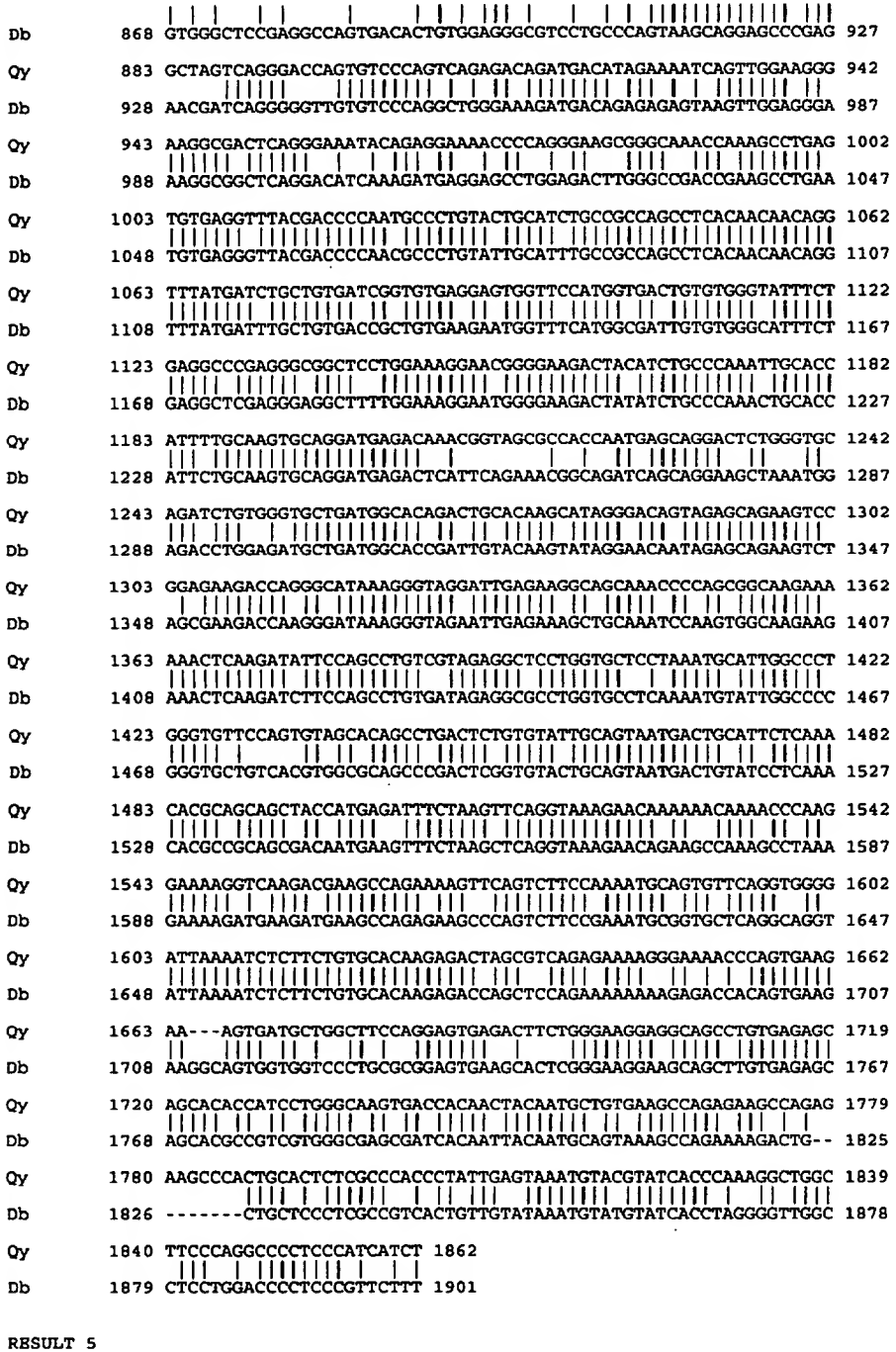


```
Qy       1689 GACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGCACACCATCCTGGGCAAGTGACCACAA 1748
Db       1770 AGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAA 1829

Qy       1749 CTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAGCCCACTGCACTCTCGCCCACCCTATT 1808
Db       1830 TTACAATGCAGTAAAGCCAGAAAAGACTG---------CTGCTCCCTCGCCGTCACTGTT 1880

Qy       1809 GAGTAAATGTACGTATCACCCAAAGGCTGGCTTCCCAGGCCCCTCCCATCATCT-----G 1863
Db       1881 GTATAAATGTATGTATCACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGAT 1940

Qy       1864 GGTGGCTGCCTGGGGCTGTCTAGGACCAG-----AGTCCTGGGTGTTCTGGTGCTGATAG 1918
Db       1941 AGCCATCCCCTGGGCCTGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAG 2000

Qy       1919 TAGCCAGCAGC----TCACTGCCAGCCAGAAGCAGATACCAAGATGCCTCTGGACCCCAG 1974
Db       2001 ACACCGGCTGCACCATCAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCT 2060

Qy       1975 GTGTTCCTGCCTAGCCTGTGGAGCCTCTCTGGGTGGTTCCTAAAGAGCTGTGTAGGCCTC 2034
Db       2061 GTGTCCATCCCTAGTCTGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCAC 2120

Qy       2035 ATGT 2038
Db       2121 TGGT 2124
```

RESULT 4
ABV28029
ID ABV28029 standard; cDNA; 2085 BP.
XX
AC ABV28029;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 28020.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 5788; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodyanamic or pharmacogenomic marker
XX
SQ Sequence 2085 BP; 520 A; 525 C; 657 G; 378 T; 0 U; 5 Other;

Query Match 33.1%; Score 949.8; DB 5; Length 2085;
Best Local Similarity 73.5%; Pred. No. 4e-286;
Matches 1384; Conservative 0; Mismatches 462; Indels 37; Gaps 12;

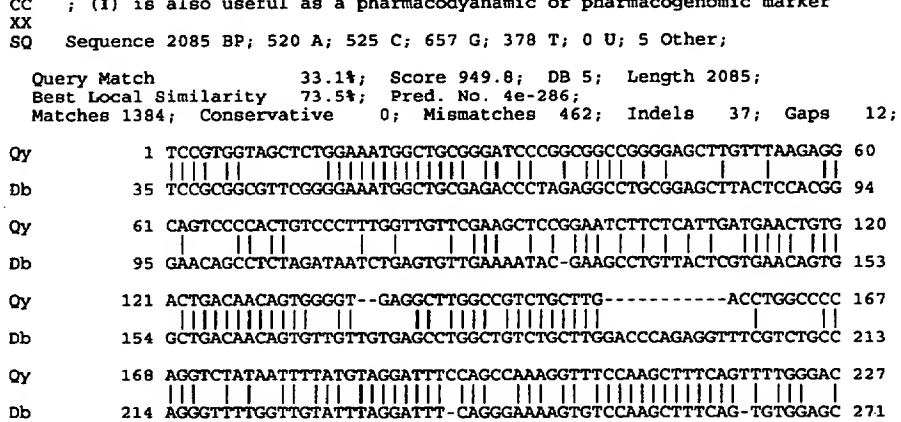
```
Qy          1 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 60
Db         35 TCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGCCTGCGGAGCTTACTCCACGG 94

Qy         61 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 120
Db         95 GAACAGCCTCTAGATAATCTGAGTGTTGAAAATAC-GAAGCCTGTTACTCGTGAACAGTG 153

Qy        121 ACTGACAACAGTGGGGT--GAGGCTTGGCCGTCTGCTTG----------ACCTGGCCCC 167
Db        154 GCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCCAGAGGTTTCGTCTGCC 213

Qy        168 AGGTCTATAATTTTATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGAC 227
Db        214 AGGGTTTTGGTTGTATTTAGGATTT-CAGGGAAAAGTGTCCAAGCTTTCAG-TGTGGAGC 271

Qy        228 AGGTATGGATGATAAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCAC 287
Db        272 AGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTAAGGCCATCAAACCCAC 331

Qy        288 CAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGG 347
Db        332 CAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTATCGCCAAGCGAGAGGG 391

Qy        348 TGCAGGAGACACGGAGGCGGACCCCAGTGAGCAGCA-----ACCACAGCAGCATAACCTC 402
Db        392 CGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCACCCCCACAGCAGCAGCTGGGCCTG 451

Qy        403 TCCCTGCGCCGCAGTGGACGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACC 462
Db        452 TCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGACC 511

Qy        463 ACGGTTCGGCGCCGAGGGAAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCC 522
Db        512 ATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGCCC 571

Qy        523 ACATCTTCCACAGTCACTGATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCT 582
Db        572 ACGTCCTGCCCCG-CACAGACGCCGAGACAGC-TCCGAGGGCAGCGTGGAAAGCGCTTCT 629

Qy        583 GAGATCAGAAGTGGCCCTGTATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAA 642
Db        630 GAGACCAGAAGCGGCCCCAGTCTGCTTCCACAGTGTGAAGGAACGACCAGCTCTTCTGAA 689

Qy        643 AAGGCAAAAGGAGGTGAAGAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACG 702
Db        690 AAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGG-CTGACC 748

Qy        703 TTGAAGGAACTTCAGAACCGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCC 762
Db        749 TTGAAAGAGCTTCAGAATCG-CTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGAGGCCC 807

Qy        763 CTGAGAGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACT 822
Db        808 CTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCGAGACT 867

Qy        823 GGGAGTGTCCAAATAGGCAGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAG 882
```

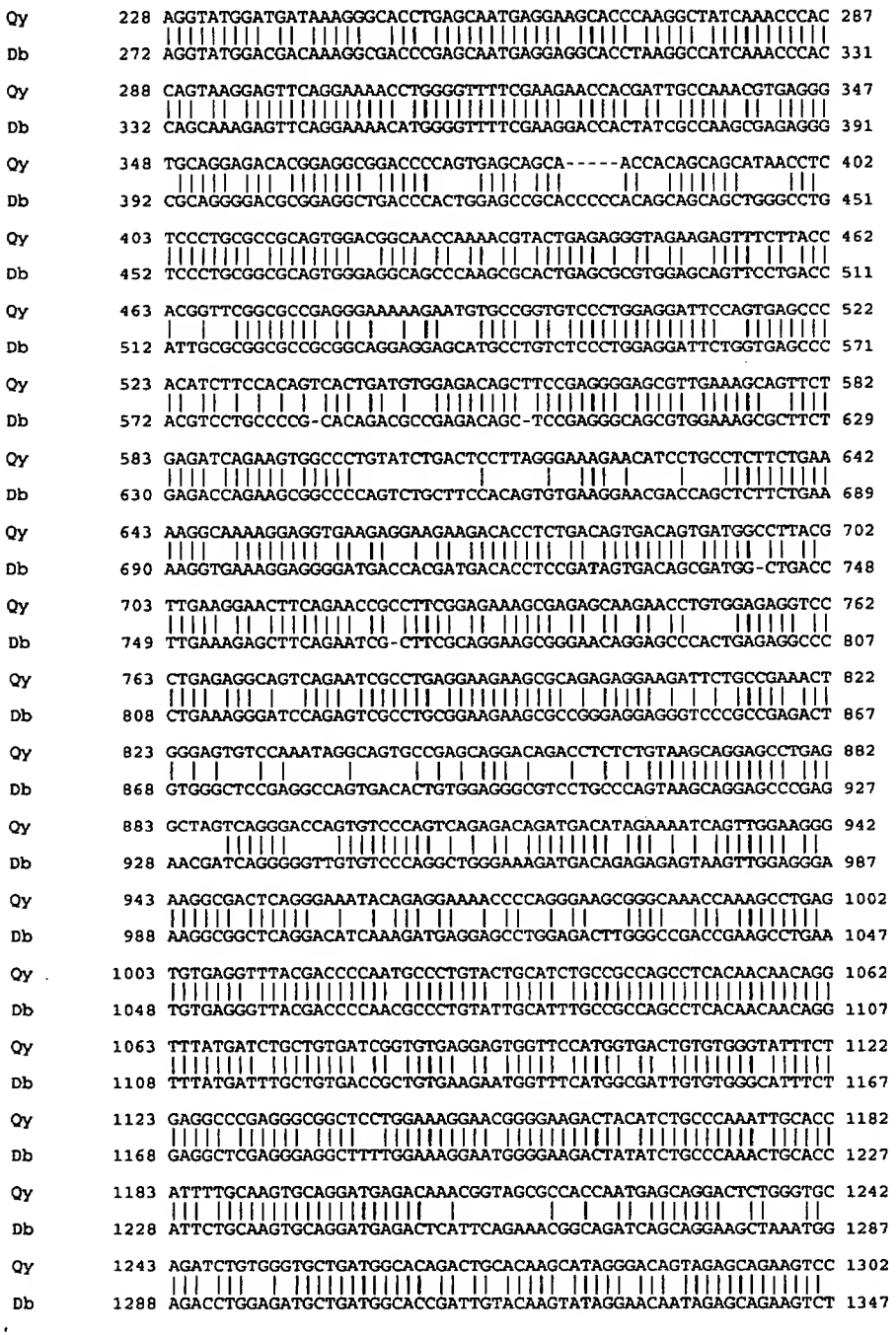


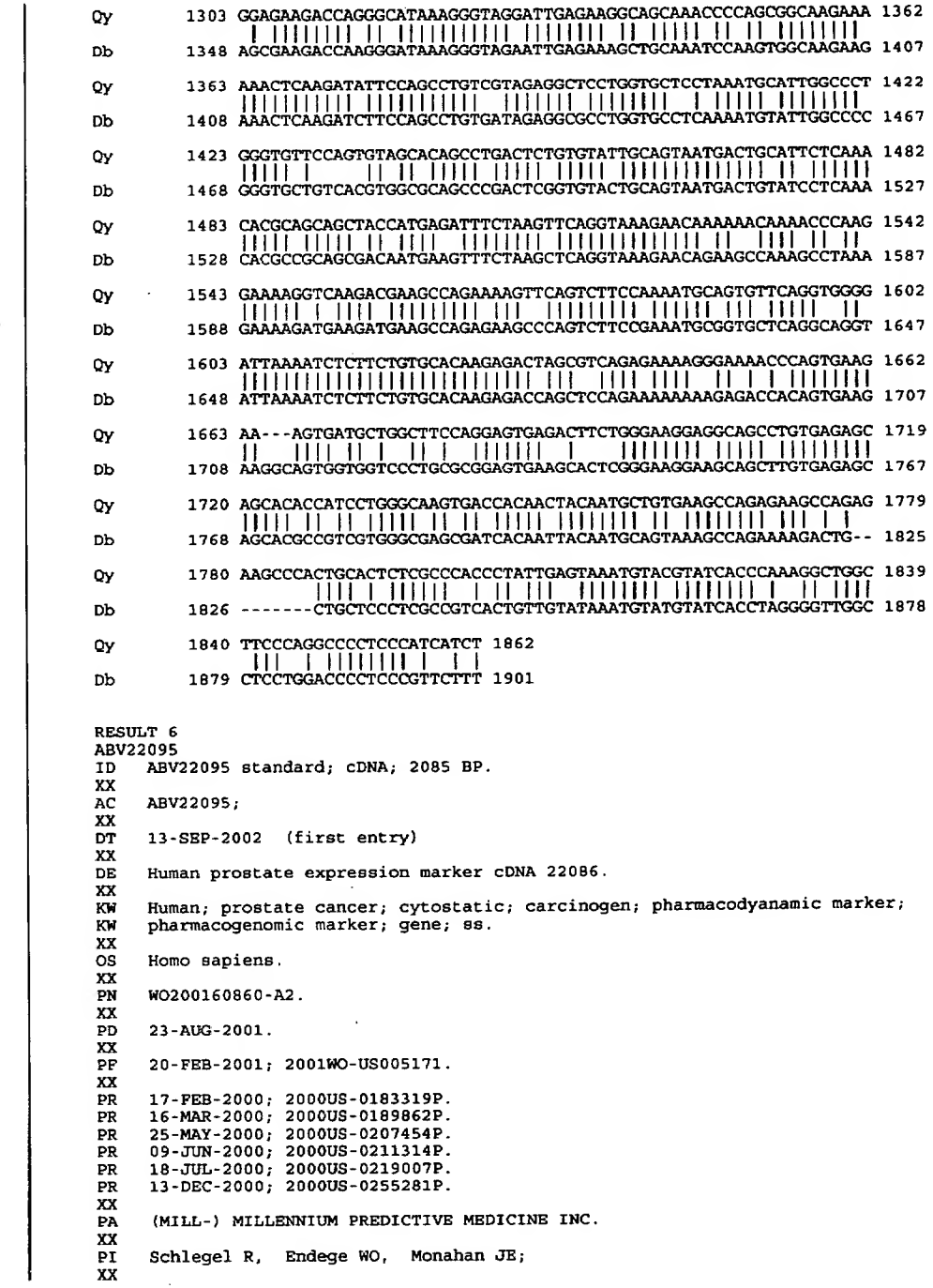
RESULT 5

ABV27935
ID ABV27935 standard; cDNA; 2085 BP.
XX
AC ABV27935;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 27926.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 5757-5758; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodyanamic or pharmacogenomic marker
XX
SQ Sequence 2085 BP; 520 A; 525 C; 657 G; 378 T; 0 U; 5 Other;

Query Match 33.1%; Score 949.8; DB 5; Length 2085;
Best Local Similarity 73.5%; Pred. No. 4e-286;
Matches 1384; Conservative 0; Mismatches 462; Indels 37; Gaps 12;







```
RESULT 6
ABV22095
ID   ABV22095 standard; cDNA; 2085 BP.
XX
AC   ABV22095;
XX
DT   13-SEP-2002  (first entry)
XX
DE   Human prostate expression marker cDNA 22086.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US005171.
XX
PR   17-FEB-2000; 2000US-0183319P.
PR   16-MAR-2000; 2000US-0189862P.
PR   25-MAY-2000; 2000US-0207454P.
PR   09-JUN-2000; 2000US-0211314P.
PR   18-JUL-2000; 2000US-0219007P.
PR   13-DEC-2000; 2000US-0255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Endege WO,  Monahan JE;
XX
```

```
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS   Claim 1; Page 3785; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for: (a) assessing whether
CC   a patient is afflicted with prostate cancer; (b) monitoring the
CC   progression of prostate cancer in a patient; (c) assessing the efficacy
CC   of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC   the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC   determining whether prostate cancer has metastasized in a patient; (h)
CC   assessing the aggressiveness or indolence of prostate cancer in a patient
CC   ; (I) is also useful as a pharmacodyanamic or pharmacogenomic marker
XX
SQ   Sequence 2085 BP; 520 A; 525 C; 657 G; 378 T; 0 U; 5 Other;

  Query Match             33.1%;  Score 949.8;  DB 5;  Length 2085;
  Best Local Similarity   73.5%;  Pred. No. 4e-286;
  Matches 1384;  Conservative    0;  Mismatches  462;  Indels   37;  Gaps   12;

Qy          1 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 60
Db         35 TCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGCCTGCGGAGCTTACTCCACGG 94

Qy         61 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 120
Db         95 GAACAGCCTCTAGATAATCTGAGTGTTGAAAATAC-GAAGCCTGTTACTCGTGAACAGTG 153

Qy        121 ACTGACAACAGTGGGGT--GAGGCTTGGCCGTCTGCTTG-----------ACCTGGCCCC 167
Db        154 GCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCCAGAGGTTTCGTCTGCC 213

Qy        168 AGGTCTATAATTTTATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGAC 227
Db        214 AGGGTTTTGGTTGTATTTAGGATTT-CAGGGAAAAGTGTCCAAGCTTTCAG-TGTGGAGC 271

Qy        228 AGGTATGGATGATAAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCAC 287
Db        272 AGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTAAGGCCATCAAACCCAC 331

Qy        288 CAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGG 347
Db        332 CAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTATCGCCAAGCGAGAGGG 391

Qy        348 TGCAGGAGACACGGAGGCGGACCCCAGTGAGCAGCA-----ACCACAGCAGCATAACCTC 402
Db        392 CGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCACCCCCACAGCAGCAGCTGGGCCTG 451

Qy        403 TCCCTGCGCCGCAGTGGACGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACC 462
Db        452 TCCCTGCGGCGCAGTGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGACC 511

Qy        463 ACGGTTCGGCGCCGAGGGAAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCC 522
Db        512 ATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGCCC 571

Qy        523 ACATCTTCCACAGTCACTGATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCT 582
Db        572 ACGTCCTGCCCCG-CACAGACGCCGAGACAGC-TCCGAGGGCAGCGTGGAAAGCGCTTCT 629

Qy        583 GAGATCAGAAGTGGCCCTGTATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAA 642
Db        630 GAGACCAGAAGCGGCCCCAGTCTGCTTCCACAGTGTGAAGGAACGACCAGCTCTTCTGAA 689

Qy        643 AAGGCAAAAGGAGGTGAAGAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACG 702
Db        690 AAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGG-CTGACC 748

Qy        703 TTGAAGGAACTTCAGAACCGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCC 762
Db        749 TTGAAAGAGCTTCAGAATCG-CTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGAGGCCC 807

Qy        763 CTGAGAGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACT 822
Db        808 CTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCGAGACT 867

Qy        823 GGGAGTGTCCAAATAGGCAGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAG 882
Db        868 GTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGCCCGAG 927

Qy        883 GCTAGTCAGGGACCAGTGTCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGG 942
Db        928 AACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGGAGGGA 987

Qy        943 AAGGCGACTCAGGGAAATACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAG 1002
Db        988 AAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACTTGGGCCGACCGAAGCCTGAA 1047

Qy       1003 TGTGAGGTTTACGACCCCAATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGG 1062
Db       1048 TGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACAACAGG 1107

Qy       1063 TTTATGATCTGCTGTGATCGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCT 1122
Db       1108 TTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCATTTCT 1167

Qy       1123 GAGGCCCGAGGGCGGCTCCTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTGCACC 1182
Db       1168 GAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACTGCACC 1227

Qy       1183 ATTTTGCAAGTGCAGGATGAGACAAACGGTAGCGCCACCAATGAGCAGGACTCTGGGTGC 1242
Db       1228 ATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTAAATGG 1287

Qy       1243 AGATCTGTGGGTGCTGATGGCACAGACTGCACAAGCATAGGACAGTAGAGCAGAAGTCC 1302
Db       1288 AGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAGGAACAATAGAGCAGAAGTCT 1347

Qy       1303 GGAGAAGACCAGGGCATAAAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCAAGAAA 1362
Db       1348 AGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTGCAAATCCAAGTGGCAAGAAG 1407

Qy       1363 AAACTCAAGATATTCCAGCCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTGGCCCT 1422
Db       1408 AAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGGTGCCTCAAAATGTATTGGCCCC 1467

Qy       1423 GGGTGTTCCAGTGTAGCACAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTCTCAAA 1482
Db       1468 GGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTGCAGTAATGACTGTATCCTCAAA 1527

Qy       1483 CACGCAGCAGCTACCATGAGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAACCCAAG 1542
Db       1528 CACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGCCTAAA 1587

Qy       1543 GAAAAGGTCAAGACGAAGCCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTGGGG 1602
Db       1588 GAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGGCAGGT 1647

Qy       1603 ATTAAAATCTCTTCTGTGCACAAGAGACTAGCGTCAGAGAAAAGGGAAAACCCAGTGAAG 1662
Db       1648 ATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAAAAAAAGAGACCACAGTGAAG 1707

Qy       1663 AA---AGTGATGCTGGCTTCCAGGAGTGAGACTTCTGGGAAGGAGGCAGCCTGTGAGAGC 1719
Db       1708 AAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTGAGAGC 1767

Qy       1720 AGCACACCATCCTGGGCAAGTGACCACAACTACAATGCTGTGAAGCCAGAGAAGCCAGAG 1779
Db       1768 AGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAGTAAAGCCAGAAAAGACTG-- 1825
```

```
Qy       1780 AAGCCCACTGCACTCTCGCCCACCCTATTGAGTAAATGTACGTATCACCCAAAGGCTGGC 1839
              |||| | ||||||  | || |||  |||||||| ||||||||  |  || ||||
Db       1826 -------CTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACCTAGGGGTTGGC 1878

Qy       1840 TTCCCAGGCCCCTCCCATCATCT 1862
                |||  | |||||||| |  | |
Db       1879 CTCCTGGACCCCTCCCGTTCTTT 1901


RESULT 7
ABV22192
ID   ABV22192 standard; cDNA; 2085 BP.
XX
AC   ABV22192;
XX
DT   13-SEP-2002  (first entry)
XX
DE   Human prostate expression marker cDNA 22183.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US005171.
XX
PR   17-FEB-2000; 2000US-0183319P.
PR   16-MAR-2000; 2000US-0189862P.
PR   25-MAY-2000; 2000US-0207454P.
PR   09-JUN-2000; 2000US-0211314P.
PR   18-JUL-2000; 2000US-0219007P.
PR   13-DEC-2000; 2000US-0255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Endege WO,  Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS   Claim 1; Page 3816; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for: (a) assessing whether
CC   a patient is afflicted with prostate cancer; (b) monitoring the
CC   progression of prostate cancer in a patient; (c) assessing the efficacy
CC   of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC   the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC   determining whether prostate cancer has metastasized in a patient; (h)
CC   assessing the aggressiveness or indolence of prostate cancer in a patient
CC   ; (I) is also useful as a pharmacodyanamic or pharmacogenomic marker
XX
SQ   Sequence 2085 BP; 520 A; 526 C; 656 G; 378 T; 0 U; 5 Other;

  Query Match             33.1%;  Score 948.2;  DB 5;  Length 2085;
  Best Local Similarity   73.4%;  Pred. No. 1.3e-285;
  Matches 1383;  Conservative    0;  Mismatches  463;  Indels   37;  Gaps   12;

Qy          1 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 60
              |||| ||       |||||||||||||| || ||  | |||| | |    |    ||
Db         35 TCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGCCTGCGGAGCTTACTCCACGG 94

Qy         61 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 120
              |    || ||       | |    | ||| | | ||| | | | | | |||| |||
Db         95 GAACAGCCTCTAGATAATCTGAGTGTTGAAAATAC-GAAGCCTGTTACTCGTGAACAGTG 153

Qy        121 ACTGACAACAGTGGGGT--GAGGCTTGGCCGTCTGCTTG----------ACCTGGCCCC 167
              |||||||||||  ||      || |||| |||||||||           |     ||
Db        154 GCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCCAGAGGTTTCGTCTGCC 213

Qy        168 AGGTCTATAATTTTATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGAC 227
              |||  | |  || ||| |||||||| |||  ||| || ||||||||||||| | |||  |
Db        214 AGGGTTTTGGTTGTATTTAGGATTT-CAGGGAAAAGTGTCCAAGCTTTCAG-TGTGGAGC 271

Qy        228 AGGTATGGATGATAAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCAC 287
              |||||||| || ||||| ||| |||||||||||| ||||| ||||| ||||||||||
Db        272 AGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTAAGGCCATCAAACCCAC 331

Qy        288 CAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGG 347
              ||| || |||||||||||||| |||||||||||||| ||||| || ||||| || |||||
Db        332 CAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTATCGCCAAGCGAGAGGG 391

Qy        348 TGCAGGAGACACGGAGGCGGACCCCAGTGAGCAGCA-----ACCACAGCAGCATAACCTC 402
              ||||| ||| ||||||| |||||    |||| |||     ||  |||||||    |||
Db        392 CGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCACCCCCACAGCAGCAGCTGGGCCTG 451

Qy        403 TCCCTGCGCCGCAGTGGACGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACC 462
              |||||||| ||||||||  |||| || || || |||||| | || ||  |||| || |||
Db        452 TCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGACC 511

Qy        463 ACGGTTCGGCGCCGAGGGAAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCC 522
              |  | |||||||| || | | ||    |||| || ||||||||||||||  ||||||||
Db        512 ATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGCCC 571

Qy        523 ACATCTTCCACAGTCACTGATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCT 582
              || || | | | | ||| || |  |||||||| |||||||| ||||| ||||||  ||||
Db        572 ACGTCCTGCCCCG-CACAGACGCCGAGACAGC-TCCGAGGGCAGCGTGGAAAGCGCTTCT 629

Qy        583 GAGATCAGAAGTGGCCCTGTATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAA 642
              |||| |||||| |||||          |     | |||  |    |  ||||||||||
Db        630 GAGACCAGAAGCGGCCCCAGTCTGCTTCCACAGTGTGAAGGAACGACCAGCTCTTCTGAA 689

Qy        643 AAGGCAAAAGGAGGTGAAGAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACG 702
              |||| |||||||| || ||  | || |||||||| || |||||||| ||||| || ||
Db        690 AAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGG-CTGACC 748

Qy        703 TTGAAGGAACTTCAGAACCGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCC 762
              ||||| || |||||||| || ||||| || ||||| || || || ||     |||||| ||
Db        749 TTGAAAGAGCTTCAGAATCG-CTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGAGGCCC 807

Qy        763 CTGAGAGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACT 822
              |||| ||| |  |||| |||||||| |||||||||| | ||||| | | ||||| |||
Db        808 CTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCGAGACT 867

Qy        823 GGGAGTGTCCAAATAGGCAGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAG 882
              | | |  | |    |    | | | ||| |  |  | | ||||||||||||| |||
Db        868 GTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGCCCGAG 927

Qy        883 GCTAGTCAGGGACCAGTGTCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGG 942
              ||||||    |||||||||| | | || |||||||| ||| | | |||||||  ||
Db        928 AACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGGAGGGA 987

Qy        943 AAGGCGACTCAGGGAAATACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAG 1002
              ||||||  ||||||    | ||| ||  | ||  |  ||||  ||| ||||||||
Db        988 AAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACTTGGGCCGACCGAAGCCTGAA 1047

Qy       1003 TGTGAGGTTTACGACCCCAATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGG 1062
              ||||||| |||||||||||| |||||||| |||||  |||||||||||||||||||||||
Db       1048 TGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACAACAGG 1107

Qy       1063 TTTATGATCTGCTGTGATCGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCT 1122
              |||||||| |||||||| || ||||| || ||||| |||||  || |||||||  ||||||
Db       1108 TTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCATTTCT 1167

Qy       1123 GAGGCCCGAGGGCGGCTCCTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTGCACC 1182
```

```
Db       1168 GAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACTGCACC 1227

Qy       1183 ATTTTGCAAGTGCAGGATGAGACAAACGGTAGCGCCACCAATGAGCAGGACTCTGGGTGC 1242

Db       1228 ATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTAAATGG 1287

Qy       1243 AGATCTGTGGGTGCTGATGGCACAGACTGCACAAGCATAGGGACAGTAGAGCAGAAGTCC 1302

Db       1288 AGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAGGAACAATAGAGCAGAAGTCT 1347

Qy       1303 GGAGAAGACCAGGGCATAAAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCAAGAAA 1362

Db       1348 AGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTGCAAATCCAAGTGGCAAGAAG 1407

Qy       1363 AAACTCAAGATATTCCAGCCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTGGCCCT 1422

Db       1408 AAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGGTGCCTCAAAATGTATTGGCCCC 1467

Qy       1423 GGGTGTTCCAGTGTAGCACAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTCTCAAA 1482

Db       1468 GGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTGCAGTAATGACTGTATCCTCAAA 1527

Qy       1483 CACGCAGCAGCTACCATGAGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAACCCAAG 1542

Db       1528 CACGCCGCACCGACAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGCCTAAA 1587

Qy       1543 GAAAAGGTCAAGACGAAGCCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTGGGG 1602

Db       1588 GAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGGCAGGT 1647

Qy       1603 ATTAAAATCTCTTCTGTGCACAAGAGACTAGCGTCAGAGAAAAGGGAAAACCCAGTGAAG 1662

Db       1648 ATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAAAAAAAGAGACCACAGTGAAG 1707

Qy       1663 AA---AGTGATGCTGGCTTCCAGGAGTGAGACTTCTGGGAAGGAGGCAGCCTGTGAGAGC 1719

Db       1708 AAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTGAGAGC 1767

Qy       1720 AGCACACCATCCTGGGCAAGTGACCACAACTACAATGCTGTGAAGCCAGAGAAGCCAGAG 1779

Db       1768 AGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAGTAAAGCCAGAAAAGACTG-- 1825

Qy       1780 AAGCCCACTGCACTCTCGCCCACCCTATTGAGTAAATGTACGTATCACCCAAAGGCTGGC 1839

Db       1826 -------CTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACCTAGGGGTTGGC 1878

Qy       1840 TTCCCAGGCCCCTCCCATCATCT 1862

Db       1879 CTCCTGGACCCCTCCCGTTCTTT 1901


RESULT 8
AAK53352
ID   AAK53352 standard; cDNA; 7290 BP.
XX
AC   AAK53352;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 2881.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US004098.
XX
PR   03-FEB-2000; 2000US-00496914.
PR   27-APR-2000; 2000US-00560875.
PR   20-JUN-2000; 2000US-00598075.
PR   19-JUL-2000; 2000US-00620325.
PR   01-SEP-2000; 2000US-00654936.
PR   15-SEP-2000; 2000US-00663561.
PR   20-OCT-2000; 2000US-00693325.
PR   30-NOV-2000; 2000US-00728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI   Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI   Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   P-PSDB; AAM80219.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT   in diagnosis and gene therapy.
XX
PS   Claim 1; Page 5053-5055; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC   (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC   sequence listing were missing at the time of publication
XX
SQ   Sequence 7290 BP; 1762 A; 1953 C; 1975 G; 1600 T; 0 U; 0 Other;

  Query Match             25.1%;  Score 719.2;  DB 4;  Length 7290;
  Best Local Similarity   70.7%;  Pred. No. 2e-213;
  Matches 1076;  Conservative    0;  Mismatches  323;  Indels  123;  Gaps    4;

Qy         416 GTGGACGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCC 475

Db          64 GTGGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCC 123

Qy         476 GAGGGAAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAG 535

Db         124 GCGGCAGGAGGAGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCG 183

Qy         536 TCACTGATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTG 595

Db         184 CCACAGACGCCGAGACAGCCTCCGAGGGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCG 243

Qy         596 GCCCTGTATCTGACTCCTTA---GGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAG 652

Db         244 GCCCCCAGTCTGCTTCCACAGCTGTGAAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAG 303

Qy         653 GAGGTGAAGAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGAAC 712

Db         304 GAGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGGCCTGACCTTGAAAGAGC 363

Qy         713 TTCAGAACCGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCA 772

Db         364 TTCAGAATCGCCTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGA 423

Qy         773 GTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCC 832

Db         424 TCCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCG 483
```

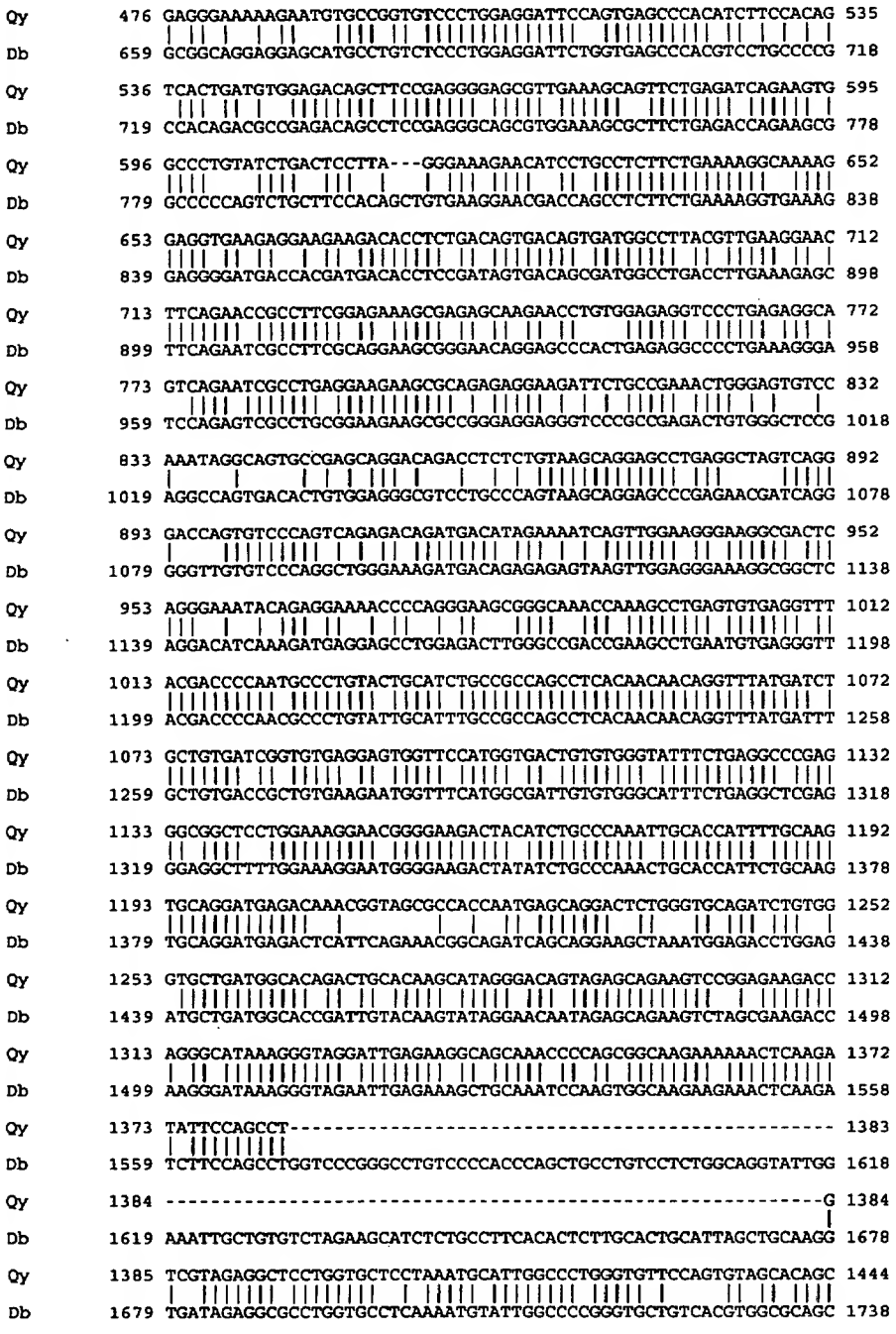
```
Qy       833 AAATAGGCAGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGG 892
Db       484 AGGCCAGTGACACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGG 543

Qy       893 GACCAGTGTCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTC 952
Db       544 GGGTTGTGTCCCAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTC 603

Qy       953 AGGGAAATACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTT 1012
Db       604 AGGACATCAAAGATGAGGAGCCTGGAGACTTGGGCCGACCGAAGCCTGAATGTGAGGGTT 663

Qy      1013 ACGACCCCAATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCT 1072
Db       664 ACGACCCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTT 723

Qy      1073 GCTGTGATCGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAG 1132
Db       724 GCTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAG 783

Qy      1133 GGCGGCTCCTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTGCACCATTTTGCAAG 1192
Db       784 GGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACTGCACCATTCTGCAAG 843

Qy      1193 TGCAGGATGAGACAAACGGTAGCGCCACCAATGAGCAGGACTCTGGGTGCAGATCTGTGG 1252
Db       844 TGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAG 903

Qy      1253 GTGCTGATGGCACAGACTGCACAAGCATAGGGACAGTAGAGCAGAAGTCCGGAGAAGACC 1312
Db       904 ATGCTGATGGCACCGATTGTACAAGTATAGGAACAATAGAGCAGAAGTCTAGCGAAGACC 963

Qy      1313 AGGGCATAAAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCAAGAAAAAACTCAAGA 1372
Db       964 AAGGGATAAAGGGTAGAATTGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGA 1023

Qy      1373 TATTCCAGCCT---------------------------------------------------- 1383
Db      1024 TCTTCCAGCCTGGTCCCGGGCCTGTCCCCACCCAGCTGCCTGTCCTCTGGCAGGTATTGG 1083

Qy      1384 ----------------------------------------------------------G 1384
Db      1084 AAATTGCTGTGTCTAGAAGCATCTCTGCCTTCACACTCTTGCACTGCATTAGCTGCAAGG 1143

Qy      1385 TCGTAGAGGCTCCTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCACAGC 1444
Db      1144 TGATAGAGGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGC 1203

Qy      1445 CTGACTCTGTGTATTGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATGAGAT 1504
Db      1204 CCGACTCGGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGT 1263

Qy      1505 TTCTAAGTTCAGGTAAAGAACAAAAAACAAAACCCAAGGAAAAGGTCAAGACGAAGCCAG 1564
Db      1264 TTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAG 1323

Qy      1565 AAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTGGGGATTAAAATCTCTTCTGTGCACA 1624
Db      1324 AGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACA 1383

Qy      1625 AGAGACTAGCGTCAGAGAAAAGGGAAAACCCAGTGAAGAA---AGTGATGCTGGCTTCCA 1681
Db      1384 AGAGACCAGCTCCAGAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGC 1443

Qy      1682 GGAGTGAGACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGCACACCATCCTGGGCAAGTG 1741
Db      1444 GGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCG 1503

Qy      1742 ACCACAACTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAGCCCACTGCACTCTCGCCCA 1801
Db      1504 ATCACAATTACAATGCAGTAAAGCCAGAAAAGACTG---------CTGCTCCCTCGCCGT 1554

Qy      1802 CCCTATTGAGTAAATGTACGTA 1823
```

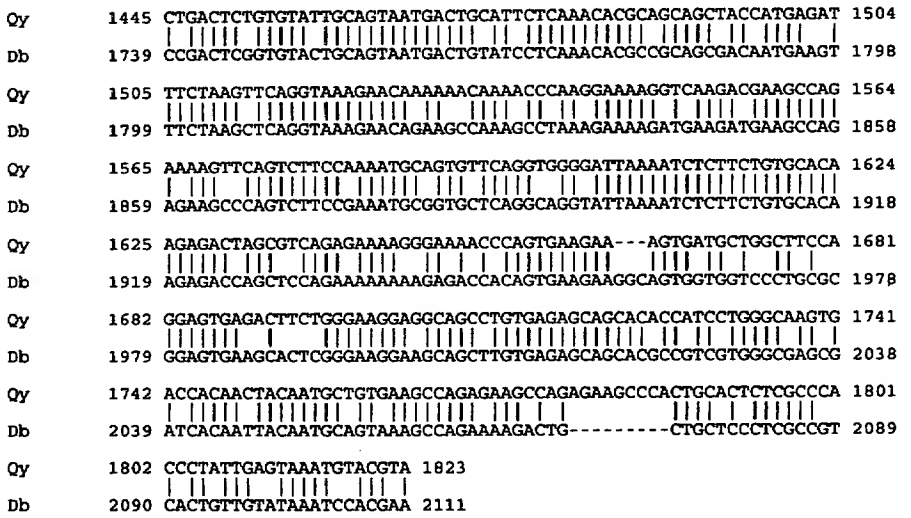
```
Db      1555 CACTGTTGTATAAATCCACGAA 1576
```

RESULT 9
AAK52368
ID AAK52368 standard; cDNA; 7838 BP.
XX
AC AAK52368;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 913.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US004098.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR P-PSDB; AAM79235.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.
XX
PS Claim 1; Page 3007-3013; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
SQ Sequence 7838 BP; 1872 A; 2123 C; 2118 G; 1725 T; 0 U; 0 Other;

Query Match 25.1%; Score 719.2; DB 4; Length 7838;
Best Local Similarity 70.7%; Pred. No. 2.1e-213;
Matches 1076; Conservative 0; Mismatches 323; Indels 123; Gaps 4;

```
Qy       416 GTGGACGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCC 475
Db       599 GTGGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCC 658
```





RESULT 10
AAS86576
ID AAS86576 standard; cDNA; 2332 BP.
XX
AC AAS86576;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #22380.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG22389.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 22380; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal

```
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 2332 BP; 511 A; 621 C; 764 G; 436 T; 0 U; 0 Other;

  Query Match           24.1%;  Score 692;  DB 5;  Length 2332;
  Best Local Similarity  71.3%;  Pred. No. 3.2e-205;
  Matches 1051;  Conservative    0;  Mismatches  335;  Indels   88;  Gaps    7;

Qy        211 GCTTTCAGTTTTGGGACAGGTATGGATGATAAAGGGCACCTGAGCAATGAGGAAGCACCC 270
Db        370 GCTTACACTTTACCGCTAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCT 429

Qy        271 AAGGCTATCAAACCCACCAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCGAAGAACCACG 330
Db        430 AAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACT 489

Qy        331 ATTGCCAAACGTGAGGGTGCAGGAGACACGGAGGCGGACCCCAGTGAGCAGCAACC---- 386
Db        490 ATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCA 549

Qy        387 --ACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGACGGCAACCAAAACGTACTGAGAGG 444
Db        550 CAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGC 609

Qy        445 GTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGGAAAAAGAATGTGCCGGTGTCCCTG 504
Db        610 GTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTG 669

Qy        505 GAGGATTCCAGTGAGCCCACATCTTCCACAGTCACTGATGTGGAGACAGCTTCCGAGGGG 564
Db        670 GAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGC 729

Qy        565 AGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCTGTATCTGACTCCTTA---GGGAAA 621
Db        730 AGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAG 789

Qy        622 GAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAAGAGGAAGAAGACACCTCTGAC 681
Db        790 GAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGAT 849

Qy        682 AGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAACCGCCTTCGGAGAAAGCGAGAG 741
Db        850 AGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAA 909

Qy        742 CAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAGA 801
Db        910 CAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGG 969

Qy        802 GAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGCAGTGCCGAGCAGGACAGACCT 861
Db        970 GAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTG 1029

Qy        862 CTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTGTCCCAGTCAGAGACAGATGAC 921
Db       1030 CCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGAC 1089

Qy        922 ATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAATACAGAGGAAAACCCCAGGGA- 980
Db       1090 AGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGAC 1149

Qy        981 AGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCCAATGCCCTGTACTGCATCTG 1040
Db       1150 TTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTG 1209

Qy       1041 CCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGATCGGTGTGAGGAGTGGTTCCA 1100
Db       1210 CCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCA 1269

Qy       1101 TGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAACGGGGAAGA 1160
Db       1270 TGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGA 1329

Qy       1161 CTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGATGAGACAAACGGTAGCGCCAC 1220
Db       1330 CTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGC 1389

Qy       1221 CAATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGATGGCACAGAC-----TGCACA 1275
Db       1390 AGATCAGCAGGAAGCTTAAATGGGAGACCTGGGAGATGCTTGATGGCACCGATTTTTACA 1449

Qy       1276 AGCATAGGGACAGT--AGAGCAGAAGTC----CGGAGAAGACCAGGGCATAAAGGGTAGG 1329
Db       1450 AGTATAGGGAACATTAGGAGCAGAAGTCTTAGCCGAAGACCCAAGGGGTTAAAGGGTAGA 1509

Qy       1330 ATTGAGAAGGCAGCAAACCCCAGCGGCAAGAAAAAACTCAAGATATTCCAGCC------- 1382
Db       1510 ATTGAGAAAGCTGCAAATCCAAGTGGCAAGAAGGAACTCAAGATCTTTCCAGCCTGTATT 1569

Qy       1383 ---------------------------------------------------------- 1382

Db       1570 GGAAATTGCTGTGTCTAGAAGCATCTCTGCCTTCACACTCTTGCACTGCATTAGCTGCAA 1629

Qy       1383 TGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCACA 1442
Db       1630 GGTGATAGAGGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCA 1689

Qy       1443 GCCTGACTCTGTGTATTGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATGAG 1502
Db       1690 GCCCGACTCGGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAA 1749

Qy       1503 ATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAACCCAAGGAAAAGGTCAAGACGAAGCC 1562
Db       1750 GTTTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCC 1809

Qy       1563 AGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAG 1596
Db       1810 AGAGAAGCCCAGTCTTCCGAAATGCGGTGCTCAG 1843

RESULT 11
AAC77933
ID   AAC77933 standard; cDNA; 1764 BP.
XX
AC   AAC77933;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human cancer associated gene sequence SEQ ID NO:327.
XX
KW   Human; cancer associated gene; cancer antigen; detection; cancer;
KW   diagnosis; cytostatic; proliferative; vulnerary; immunomodulator;
KW   antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW   antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW   dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW   vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW   immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW   allergic reaction; graft versus host disease; organ rejection;
KW   haemostatic; thrombolytic; cardiovascular disorder; infection;
KW   neurological disease; drug screening; ss.
XX
OS   Homo sapiens.
XX
```

```
PN   WO200055350-A1.
XX
PD   21-SEP-2000.
XX
PP   08-MAR-2000; 2000WO-US005882.
XX
PR   12-MAR-1999;   99US-0124270P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587533/55.
DR   P-PSDB; AAB43724.
XX
PT   Novel isolated nucleic acids comprising sequences encoding peptides
PT   useful for treating or diagnosing e.g. cancer.
XX
PS   Claim 1; Page 880-881; 2352pp; English.
XX
CC   AAC77607 to AAC78448 encode the human cancer associated proteins given in
CC   AAB43398 to AAB44239. The proteins can have activities based on the
CC   tissues and cells the genes are expressed in. Example of activities
CC   include: cytostatic; proliferative; vulnerary; immunomodulator;
CC   antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC   antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC   dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC   nootropic; vasotropic; antipsoriatic and antiangiogenic. The
CC   polynucleotides and polypeptides can be used for preventing, treating or
CC   ameliorating medical conditions and diagnosing pathological conditions.
CC   Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC   the present invention may be used to treat immune disorders by activating
CC   or inhibiting the proliferation, differentiation or mobilisation of
CC   immune cells, to treat disorders of haematopoietic cells, autoimmune
CC   disorders, allergic reactions, graft versus host disease and organ
CC   rejection, modulate haemostatic or thrombolytic activity, modulate
CC   inflammation, cancers, cardiovascular disorders, neurological disease and
CC   bacterial or viral infections. The peptides, nucleotides, antibodies,
CC   agonists and antagonists may be also be used in drug screens. AAC78449 to
CC   AAC78457 and AAB44240 represent sequences used in the exemplification of
CC   the present invention
XX
SQ   Sequence 1764 BP; 468 A; 424 C; 415 G; 453 T; 0 U; 4 Other;

  Query Match           19.5%;  Score 558;  DB 3;  Length 1764;
  Best Local Similarity  74.2%;  Pred. No. 2.8e-163;
  Matches  795;  Conservative    0;  Mismatches  250;  Indels   27;  Gaps    6;

Qy       984 GGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCCAATGCCCTGTACTGCATCTGCCG 1043
Db        32 GGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCCG 91

Qy      1044 CCAGCCTCACAACAACAGGTTTATGATCTGCTGTGATCGGTGTGAGGAGTGGTTCCATGG 1103
Db        92 CCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATGG 151

Qy      1104 TGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAACGGGGAAGACTA 1163
Db       152 CGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACTA 211

Qy      1164 CATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGATGAGACAAACGGTAGCGCCACCAA 1223
Db       212 TATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAGA 271

Qy      1224 TGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGATGGCACAGACTGCACAAGCATAGG 1283
Db       272 TCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAGG 331

Qy      1284 GACAGTAGAGCAGAAGTCCGGAGAAGACCAGGGCATAAAGGGTAGGATTGAGAAGGCAGC 1343
Db       332 AACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTGC 391

Qy      1344 AAACCCCAGCGGCAAGAAAAAACTCAAGATATTCCAGCCTGTCGTAGAGGCTCCTGGTGC 1403
Db       392 AAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGGTGC 451

Qy      1404 TCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCACAGCCTGACTCTGTGTATTGCAG 1463
Db       452 CTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCA-CCCGACTCGGTGTACTGCAG 510

Qy      1464 TAATGACTGCATTCTCAAACACGCAGCAGCTACCATGAGATTTCTAAGTTCAGGTAAAGA 1523
Db       511 TAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAAAGA 570

Qy      1524 ACAAAAAACAAAACCCAAGGAAAAGGTCAAGACGAAGCCAGAAAAGTTCAGTCTTCCAAA 1583
Db       571 ACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGAA 630

Qy      1584 ATGCAGTGTTCAGGTGGGGATTAAAATCTCTTCTGTGCACAAGAGACTAGCGTCAGAGAA 1643
Db       631 ATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAAA 690

Qy      1644 AAGGGAAAACCCAGTGAAGAA---AGTGATGCTGGCTTCCAGGAGTGAGACTTCTGGGAA 1700
Db       691 AAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGAA 750

Qy      1701 GGAGGCAGCCTGTGAGAGCAGCACACCATCCTGGGCAAGTGACCACAACTACAATGCTGT 1760
Db       751 GGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAGT 810

Qy      1761 GAAGCCAGAGAAGCCAGAGAAGCCCACTGCACTCTCGCCCACCCTATTGAGTAAATGTAC 1820
Db       811 AAAGCCAGAAAAGACTG---------CTGCTCCCTCGCCGTCACTGTTGTATAAATGTAT 861

Qy      1821 GTATCACCCAAAGGCTGGCTTCCCAGGCCCCTCCCATCATCT-----GGGTGGCTGCCTG 1875
Db       862 GTATCACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTG 921

Qy      1876 GGGCTGTCTAGGACCAG-----AGTCCTGGGTGTTCTGGTGCTGATAGTAGCCAGCAGC- 1929
Db       922 GGCCTGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCA 981

Qy      1930 ---TCACTGCCAGCCAGAAGCAGATACCAAGATGCCTCTGGACCCCAGGTGTTCCTGCCT 1986
Db       982 CCATCAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCT 1041

Qy      1987 AGCCTGTGGAGCCTCTCTGGGTGGTTCCTAAAGAGCTGTGTAGGCCTCATGT 2038
Db       1042 AGTCTGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGT 1093

RESULT 12
AAA43780
ID   AAA43780 standard; cDNA; 408 BP.
XX
AC   AAA43780;
XX
DT   21-AUG-2000  (first entry)
XX
DE   Mouse secreted expressed sequence tag SEQ ID NO:355.
XX
KW   Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;
KW   expressed sequence tag; EST; probe; chemotactic; proliferative;
KW   immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW   thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW   antiviral; antidiabetic; antiasthmatic; vulnerary; antiparkinsonian;
KW   antiulcer; osteopathic; neuroprotective; nootropic; antipsoriatic;
KW   cerebroprotective; anticonvulsant; antidepressant; gene therapy; vaccine;
KW   autoimmune disorder; multiple sclerosis; allergic condition;
KW   insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW   lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW   central nervous system disorder; Alzheimer's disease; stroke;
KW   Parkinson's disease; Huntington's disease; coagulation disorder;
KW   haemophilia; thrombosis; inflammatory disorder; Crohn's disease; tumour;
KW   infection; depression; psoriasis; ss.
XX
```

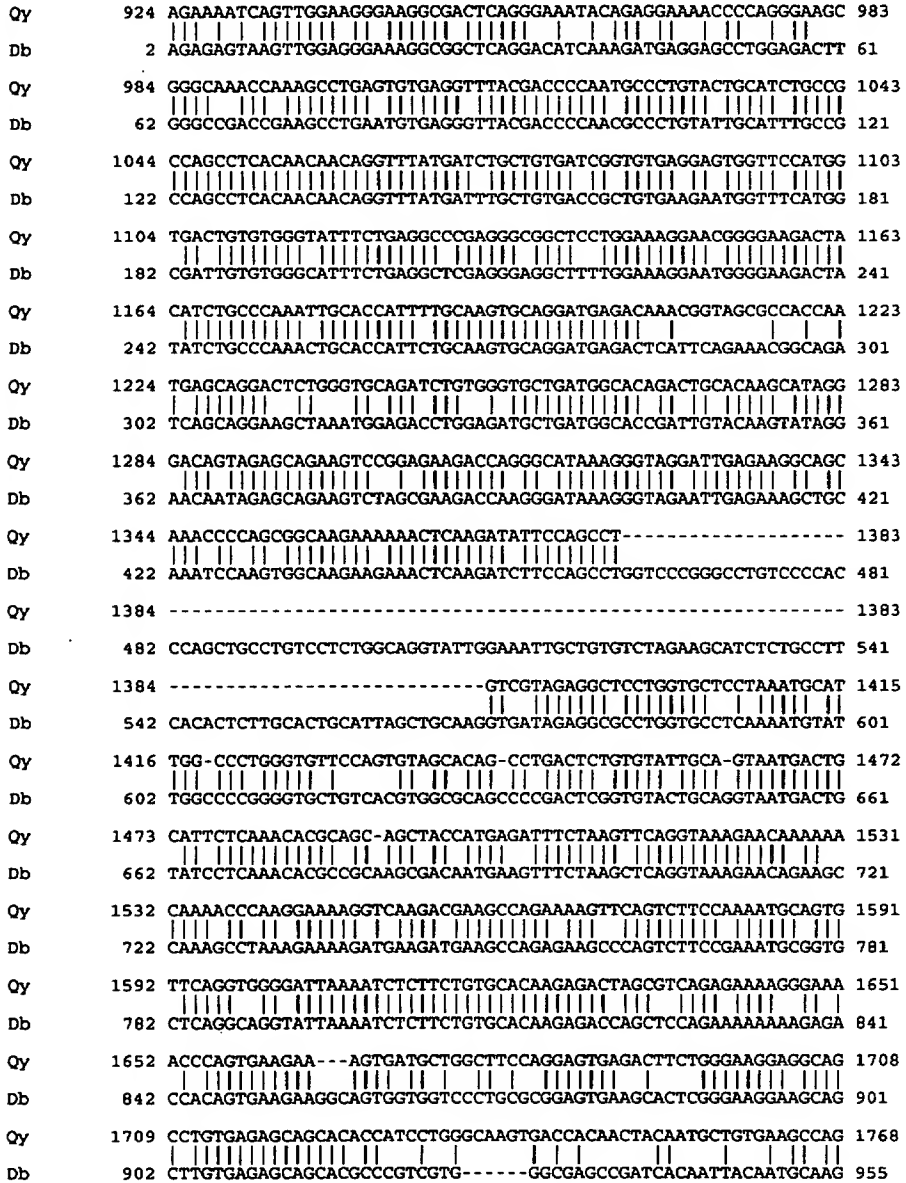
OS Mus musculus.
XX
PN WO200021991-A1.
XX
PD 20-APR-2000.
XX
PF 15-OCT-1999; 99WO-US024206.
XX
PR 15-OCT-1998; 98US-0104436P.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, Mccoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Bowman MR;
XX
DR WPI; 2000-317938/27.
XX
PT Isolated polynucleotides, and encoded proteins, comprising secreted
PT expressed sequence tags (sESTs), useful for treating various disorders
PT such as autoimmune, infectious, and central nervous system disorders.
XX
PS Claim 1; Page 296-297; 803pp; English.
XX
CC AAA43426 to AAA45925 represent specifically claimed secreted expressed
CC sequence tags (sESTs), isolated from human, mouse, chicken and rat tissue
CC sources. The sESTs can have a range of activities depending on the
CC tissues they were isolated from. The activities include: chemotactic;
CC proliferative; immunomodulatory; haematopoietic; chemokinetic; analgesic;
CC haemostatic; thrombolytic; antiinflammatory; cytostatic; antibacterial;
CC antifungal; antiviral; antidiabetic; antiasthmatic; vulnerary; antiulcer;
CC osteopathic; neuroprotective; nootropic; antiparkinsonian; antipsoriatic;
CC cerebroprotective; anticonvulsant; and antidepressant. The sESTs can be
CC used for gene therapy and in vaccines. The sESTs are useful as probes for
CC the identification and isolation of full-length cDNAs and genomic DNA
CC molecules which correspond to the sESTs. Proteins encoded by the sESTs
CC are useful in assays for determining biological activity and raising
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC in the exemplification of the present invention
XX
SQ Sequence 408 BP; 116 A; 98 C; 122 G; 72 T; 0 U; 0 Other;

Query Match 13.4%; Score 382.8; DB 3; Length 408;
Best Local Similarity 99.5%; Pred. No. 9.3e-109;
Matches 384; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      197 CCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGATAAAGGGCACCTGAGCA 256
            | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       19 CTAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGATAAAGGGCACCTGAGCA 78

Qy      257 ATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTCAGGAAAACCTGGGGTT 316
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       79 ATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTCAGGAAAACCTGGGGTT 138

Qy      317 TTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACGGAGGCGGACCCCAGTG 376
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      139 TTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACGGAGGTGGACCCCAGTG 198

Qy      377 AGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGACGGCAACCAAAACGTA 436
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      199 AGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGACGGCAACCAAAACGTA 258

Qy      437 CTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGGAAAAAGAATGTGCCGG 496
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      259 CTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGGAAAAAGAATGTGCCGG 318

Qy      497 TGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACTGATGTGGAGACAGCTT 556
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      319 TGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACTGATGTGGAGACAGCTT 378

Qy      557 CCGAGGGGAGCGTTGAAAGCAGTTCT 582
            |||||||||||||||||||||||||
Db      379 CCGAGGGGAGCGTTGAAAGCAGTTCT 404
```

RESULT 13
AAS86575
ID AAS86575 standard; cDNA; 2986 BP.
XX
AC AAS86575;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #22379.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG22388.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 22379; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2986 BP; 798 A; 796 C; 781 G; 611 T; 0 U; 0 Other;

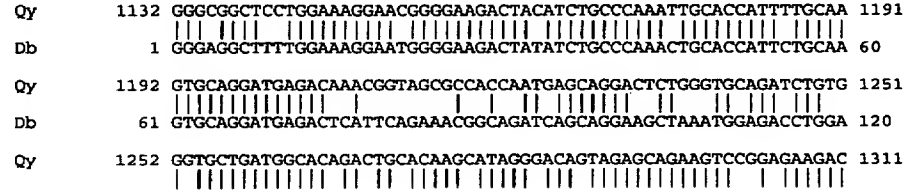
Query Match 13.1%; Score 376; DB 5; Length 2986;
Best Local Similarity 68.6%; Pred. No. 4.9e-106;
Matches 681; Conservative 0; Mismatches 190; Indels 121; Gaps 7;



RESULT 14
AAS86574
ID AAS86574 standard; cDNA; 503 BP.

XX
AC AAS86574;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #22378.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG22387.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 22378; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 503 BP; 165 A; 107 C; 138 G; 93 T; 0 U; 0 Other;

Query Match 12.1%; Score 347.8; DB 5; Length 503;
Best Local Similarity 80.7%; Pred. No. 1e-97;
Matches 406; Conservative 0; Mismatches 97; Indels 0; Gaps 0;



```
Db        121 GATGCTGATGGCACCGATTGTACAAGTATAGGAACAATAGAGCAGAAGTCTAGCGAAGAC 180

Qy       1312 CAGGGCATAAAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCAAGAAAAAACTCAAG 1371

Db        181 CAAGGGATAAAGGGTAGAATTGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAAACTCAAG 240

Qy       1372 ATATTCCAGCCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCC 1431

Db        241 ATCTTCCAGCCTGTGATAGAGGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGT 300

Qy       1432 AGTGTAGCACAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTCTCAAACACGCAGCA 1491

Db        301 CACGTGGCGCAGCCCGACTCGGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCA 360

Qy       1492 GCTACCATGAGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAACCCAAGGAAAAGGTC 1551

Db        361 GCGACAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATG 420

Qy       1552 AAGACGAAGCCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTGGGGATTAAAATC 1611

Db        421 AAGATGAAGCCAGAGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATC 480

Qy       1612 TCTTCTGTGCACAAGAGACTAGC 1634

Db        481 TCTTCTGTGCACAAGAGACCAGC 503
```

RESULT 15
ABN21248
ID ABN21248 standard; cDNA; 324 BP.
XX
AC ABN21248;
XX
DT 24-JUN-2002 (first entry)
XX
DE Human ORFX polynucleotide sequence SEQ ID NO:10973.
XX
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200192523-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US010836.
XX
PR 30-MAY-2000; 2000US-0206132P.
PR 29-AUG-2000; 2000US-0228716P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach MD;
XX
DR WPI; 2002-106308/14.
DR P-PSDB; ABP05496.
XX
PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders.
XX
PS Disclosure; SEQ ID NO 10973; 1037pp; English.
XX
CC The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage. N.B. The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 324 BP; 106 A; 72 C; 78 G; 68 T; 0 U; 0 Other;

Query Match 11.2%; Score 322.4; DB 6; Length 324;
Best Local Similarity 99.7%; Pred. No. 7e-90;
Matches 323; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1300 TCCGGAGAAGACCAGGGCATAAAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCAAG 1359

Db          1 TCCGGAGAAGACCATGGCATAAAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCAAG 60

Qy       1360 AAAAAACTCAAGATATTCCAGCCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTGGC 1419

Db         61 AAAAAACTCAAGATATTCCAGCCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTGGC 120

Qy       1420 CCTGGGTGTTCCAGTGTAGCACAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTCTC 1479

Db        121 CCTGGGTGTTCCAGTGTAGCACAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTCTC 180

Qy       1480 AAACACGCAGCAGCTACCATGAGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAACCC 1539

Db        181 AAACACGCAGCAGCTACCATGAGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAACCC 240

Qy       1540 AAGGAAAAGGTCAAGACGAAGCCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTG 1599

Db        241 AAGGAAAAGGTCAAGACGAAGCCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTG 300

Qy       1600 GGGATTAAAATCTCTTCTGTGCAC 1623

Db        301 GGGATTAAAATCTCTTCTGTGCAC 324
```

Search completed: April 23, 2004, 08:28:30
Job time : 1074.35 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 23, 2004, 04:42:12 ; Search time 10916.3 Seconds
(without alignments)
11383.418 Million cell updates/sec

Title: US-09-787-016A-2
Perfect score: 2867
Sequence: 1 tccgtggtagctctggaaat.....tttgtttcctttgtgctggg 2867

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2867	100.0	2867	6	BD244713	BD244713 Genes enc
2	2867	100.0	2867	6	AX023370	AX023370 Sequence
3	2602	90.8	2602	10	MMU238332	AJ238332 Mus muscu
4	1770.4	61.8	4906	10	AK129117	AK129117 Mus muscu
c 5	1241.2	43.3	104663	10	AL732560	AL732560 Mouse DNA
6	1023.6	35.7	2610	6	BD244712	BD244712 Genes enc
7	1023.6	35.7	2610	6	AX023369	AX023369 Sequence
8	1016.2	35.4	2767	9	BC004237	BC004237 Homo sapi
9	1013.6	35.4	2725	9	BC014489	BC014489 Homo sapi
10	1013	35.3	2772	6	AX878224	AX878224 Sequence
11	1013	35.3	2772	6	BD157094	BD157094 Primer fo
12	1013	35.3	2772	9	AK002127	AK002127 Homo sapi
13	985	34.4	2407	9	BC000770	BC000770 Homo sapi
14	853.4	29.8	221547	2	AC119716	AC119716 Rattus no
15	748.2	26.1	272404	2	AC108337	AC108337 Rattus no
16	703.4	24.5	69252	2	AC101519	AC101519 Mus muscu
17	449.8	15.7	6692	9	AB002331	AB002331 Human mRN
c 18	365.4	12.7	28519	9	HSJ563E14	AL117379 Human DNA
c 19	324.4	11.3	69252	2	AC101519	AC101519 Mus muscu
c 20	230.4	8.0	160241	9	HS885L7	AL035669 Human DNA
21	230.2	8.0	2428	5	BC060442	BC060442 Xenopus l
22	209.8	7.3	681	6	AX869935	AX869935 Sequence
23	209.8	7.3	681	6	BD149997	BD149997 Primer fo
c 24	184.6	6.4	812	11	BV013961	BV013961 S212P6852
c 25	108.6	3.8	198133	2	BX296530	BX296530 Danio rer
c 26	77.8	2.7	7218	6	I66494	I66494 Sequence 14
c 27	70.2	2.4	61204	2	AC017132	AC017132 Drosophil
28	70.2	2.4	161601	3	AC007594	AC007594 Drosophil
29	70.2	2.4	225655	3	AE003695	AE003695 Drosophil
30	69.6	2.4	2161	3	AK114522	AK114522 Ciona int
31	57.8	2.0	2308	9	BC015733	BC015733 Homo sapi
32	57.8	2.0	2320	6	AX086854	AX086854 Sequence
33	57.8	2.0	2320	9	HSM801830	AL136862 Homo sapi
34	57.8	2.0	2331	9	BC014940	BC014940 Homo sapi
35	57.8	2.0	2387	10	BC030938	BC030938 Mus muscu
36	57.8	2.0	2452	9	AF149758	AF149758 Homo sapi
37	57.8	2.0	2487	9	AB031069	AB031069 Homo sapi
38	57.8	2.0	2509	10	AY099096	AY099096 Mus muscu
39	57.8	2.0	2967	9	HSM807333	BX647189 Homo sapi
40	56.8	2.0	1310	3	AY089443	AY089443 Drosophil
41	56	2.0	10732	6	E32986	E32986 Gene encodi
42	55.4	1.9	2888	9	AK055678	AK055678 Homo sapi
43	55.4	1.9	5975	6	AX329603	AX329603 Sequence
44	55.4	1.9	6256	6	AR338778	AR338778 Sequence
45	55.4	1.9	6936	9	D87685	D87685 Human mRNA

ALIGNMENTS

RESULT 1
BD244713
LOCUS BD244713 2867 bp DNA linear PAT 17-JUL-2003
DEFINITION Genes encoding for the human and murine death inducer-obliterator-1.
ACCESSION BD244713
VERSION BD244713.1 GI:33054483
KEYWORDS JP 2002526040-A/2.
SOURCE Mus sp.
ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2867)
AUTHORS Alonso,C.M., Domingo,D.G., Grandien,A., Leonardo,E. and Martinez,P.
TITLE Genes encoding for the human and murine death inducer-obliterator-1

```
  JOURNAL   Patent: JP 2002526040-A 2 20-AUG-2002;
            CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS
COMMENT     OS   Mus sp. (murine)
            PN   JP 2002526040-A/2
            PD   20-AUG-2002
            PF   10-SEP-1999 JP 2000570314
            PR   10-SEP-1998 SE    9803069-5,17-SEP-1998 US    60/100873 PI
            CARLOS MARTINEZ ALONSO,DAVID GARCIA DOMINGO,ALF GRANDIEN, PI
            ESTHER LEONARDO,
            PI   PEDRO MARTINEZ
            PC   C12N15/09,A61K38/00,A61K45/00,A61K48/00,A61P3/10,A61P17/00, PC
              A61P29/00,
            PC   A61P35/00,A61P37/06,A61P43/00,C07K14/47,C07K16/18,C12N1/21, PC
              C12N5/10,
            PC   C12P21/02,C12Q1/02,C12Q1/68,G01N33/15,G01N33/50//C12P21/08, PC
              C12N15/00,
            PC   C12N5/00,A61K37/02
            CC   Genes encoding for the human and murine death inducer- CC
                 obliterator-1
            FH   Key             Location/Qualifiers
            FT   source          1. .2867
            FT                   /organism='Mus sp. (murine)'.
FEATURES            Location/Qualifiers
     source         1. .2867
                    /organism="Mus sp."
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10095"
ORIGIN

  Query Match           100.0%;  Score 2867;  DB 6;  Length 2867;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 2867;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 60

Qy         61 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 120

Qy        121 ACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTT 180

Qy        181 TATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGAT 240

Qy        241 AAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTC 300

Qy        301 AGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACG 360

Qy        361 GAGGCGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAGGCGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGA 420

Qy        421 CGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGG 480

Qy        481 AAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACT 540

Qy        541 GATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCT 600

Qy        601 GTATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAA 660

Qy        661 GAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAAC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAAC 720

Qy        721 CGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAAT 780

Qy        781 CGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGC 840

Qy        841 AGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTG 900

Qy        901 TCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAAT 960

Qy        961 ACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCC 1020

Qy       1021 AATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGAT 1080

Qy       1081 CGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTC 1140

Qy       1141 CTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGAT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGAT 1200

Qy       1201 GAGACAAACGGTAGCGCCACCAATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGAT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGACAAACGGTAGCGCCACCAATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGAT 1260

Qy       1261 GGCACAGACTGCACAAGCATAGGGACAGTAGAGCAGAAGTCCGGAGAAGACCAGGGCATA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GGCACAGACTGCACAAGCATAGGGACAGTAGAGCAGAAGTCCGGAGAAGACCAGGGCATA 1320

Qy       1321 AAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCAAGAAAAAACTCAAGATATTCCAG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCAAGAAAAAACTCAAGATATTCCAG 1380

Qy       1381 CCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCA 1440

Qy       1441 CAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATG 1500

Qy       1501 AGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAACCCAAGGAAAAGGTCAAGACGAAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAACCCAAGGAAAAGGTCAAGACGAAG 1560

Qy       1561 CCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTGGGGATTAAAATCTCTTCTGTG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTGGGGATTAAAATCTCTTCTGTG 1620

Qy       1621 CACAAGAGACTAGCGTCAGAGAAAAGGGAAAACCCAGTGAAGAAAGTGATGCTGGCTTCC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CACAAGAGACTAGCGTCAGAGAAAAGGGAAAACCCAGTGAAGAAAGTGATGCTGGCTTCC 1680

Qy       1681 AGGAGTGAGACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGCACACCATCCTGGGCAAGT 1740
```

```
Db      1681 AGGAGTGAGACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGCACACCATCCTGGGCAAGT 1740
Qy      1741 GACCACAACTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAGCCCACTGCACTCTCGCCC 1800
Db      1741 GACCACAACTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAGCCCACTGCACTCTCGCCC 1800
Qy      1801 ACCCTATTGAGTAAATGTACGTATCACCCAAAGGCTGGCTTCCCAGGCCCCTCCCATCAT 1860
Db      1801 ACCCTATTGAGTAAATGTACGTATCACCCAAAGGCTGGCTTCCCAGGCCCCTCCCATCAT 1860
Qy      1861 CTGGGTGGCTGCCTGGGGCTGTCTAGGACCAGAGTCCTGGGTGTTCTGGTGCTGATAGTA 1920
Db      1861 CTGGGTGGCTGCCTGGGGCTGTCTAGGACCAGAGTCCTGGGTGTTCTGGTGCTGATAGTA 1920
Qy      1921 GCCAGCAGCTCACTGCCAGCCAGAAGCAGATACCAAGATGCCTCTGGACCCCAGGTGTTC 1980
Db      1921 GCCAGCAGCTCACTGCCAGCCAGAAGCAGATACCAAGATGCCTCTGGACCCCAGGTGTTC 1980
Qy      1981 CTGCCTAGCCTGTGGAGCCTCTCTGGGTGGTTCCTAAAGAGCTGTGTAGGCCTCATGTTG 2040
Db      1981 CTGCCTAGCCTGTGGAGCCTCTCTGGGTGGTTCCTAAAGAGCTGTGTAGGCCTCATGTTG 2040
Qy      2041 GAGGCAATTTCTTATTTCAGTTTTAGGCCTTGGTAAGTATTTGAACTGCTCTCAACAAGA 2100
Db      2041 GAGGCAATTTCTTATTTCAGTTTTAGGCCTTGGTAAGTATTTGAACTGCTCTCAACAAGA 2100
Qy      2101 TGGGACAACATCAGTGCACTGTTCAGACTCAGTTCAGACTTGAGTTCCTCACAGGACAGC 2160
Db      2101 TGGGACAACATCAGTGCACTGTTCAGACTCAGTTCAGACTTGAGTTCCTCACAGGACAGC 2160
Qy      2161 CCAGCACAAACCATGGAGTTTTCCCACGTTACCTATTTTTCTTCAACATGGTCACCACTG 2220
Db      2161 CCAGCACAAACCATGGAGTTTTCCCACGTTACCTATTTTTCTTCAACATGGTCACCACTG 2220
Qy      2221 CTCTTCTATAGAAAAGTGATTTTTTTCATGAGTTAGGGCATTTTTGAAAGGACAGGTAGA 2280
Db      2221 CTCTTCTATAGAAAAGTGATTTTTTTCATGAGTTAGGGCATTTTTGAAAGGACAGGTAGA 2280
Qy      2281 CAATTTGTGTCTATTTCACATACTACACACCTACATAAACAGGTTTGAATTTTGAAGGTC 2340
Db      2281 CAATTTGTGTCTATTTCACATACTACACACCTACATAAACAGGTTTGAATTTTGAAGGTC 2340
Qy      2341 ATCTGCTGCTCAGATCCCATGGTGAGTAATCGTGCTGACCATACACCTACAGCAGTATTC 2400
Db      2341 ATCTGCTGCTCAGATCCCATGGTGAGTAATCGTGCTGACCATACACCTACAGCAGTATTC 2400
Qy      2401 TCATCAAAGTGAGAATTGTGGGTGGTAGCTTCCAGACTCGTAGATGCCCCAGATTTCAAA 2460
Db      2401 TCATCAAAGTGAGAATTGTGGGTGGTAGCTTCCAGACTCGTAGATGCCCCAGATTTCAAA 2460
Qy      2461 TCAGTGTTTATAGTTTGAGTAAACTTTTGAAACCCAAGATTAGATCACTGAGTTTGGAGC 2520
Db      2461 TCAGTGTTTATAGTTTGAGTAAACTTTTGAAACCCAAGATTAGATCACTGAGTTTGGAGC 2520
Qy      2521 CATTCTTGTTTCACTGCTGTACAAGGTACACCTTTCTTACTTTTTGTACAGTGGTTTCAA 2580
Db      2521 CATTCTTGTTTCACTGCTGTACAAGGTACACCTTTCTTACTTTTTGTACAGTGGTTTCAA 2580
Qy      2581 GTGAATTTTCATCTGTGTTCTCATGATAGTGTTTTGTTCCATCTCTGTCCACCCTCAGCC 2640
Db      2581 GTGAATTTTCATCTGTGTTCTCATGATAGTGTTTTGTTCCATCTCTGTCCACCCTCAGCC 2640
Qy      2641 CCAGATAAGGGGGGTCTCATGGCCTAAGCTGGCCTTAAATTTTCTGTAAAGTTGGGGGTCC 2700
Db      2641 CCAGATAAGGGGGGTCTCATGGCCTAAGCTGGCCTTAAATTTTCTGTAAAGTTGGGGGTCC 2700
Qy      2701 TCCTGCCCCCACCTGTAAAGAGTGCTGAAACTTACAGGCACACACCTGATGTATGCTGTG 2760
Db      2701 TCCTGCCCCCACCTGTAAAGAGTGCTGAAACTTACAGGCACACACCTGATGTATGCTGTG 2760
Qy      2761 GGGTGTGAACTCAGAGTGTTGGACCAGCAGTCTACCAGCTGAGCTGCAGTTCTAGCCATG 2820
Db      2761 GGGTGTGAACTCAGAGTGTTGGACCAGCAGTCTACCAGCTGAGCTGCAGTTCTAGCCATG 2820
Qy      2821 GTAACTGAAACTGCTGATCTTCTTGTCTTTGTTTCCTTTGTGCTGGG 2867
Db      2821 GTAACTGAAACTGCTGATCTTCTTGTCTTTGTTTCCTTTGTGCTGGG 2867
```

RESULT 2
AX023370
LOCUS AX023370 2867 bp DNA linear PAT 15-SEP-2000
DEFINITION Sequence 2 from Patent WO0015787.
ACCESSION AX023370
VERSION AX023370.1 GI:10183782
KEYWORDS .
SOURCE Mus sp.
ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Leonardo,E., Martinez,P., Alonso,C.M., Domingo,D.G. and Grandien,A.
TITLE Genes encoding for the human and murine death in ducer-obliterator-1
JOURNAL Patent: WO 0015787-A 2 23-MAR-2000;
LEONARDO ESTHER (ES) ; MARTINEZ PEDRO (ES) ; CONSEJO SUPERIOR INVESTIGACION (ES) ; DOMINGO DAVID GARCIA (ES) ; MARTINEZ ALONSO CARLOS (ES) ; BANNERMAN DAVID GARDNER (GB) ; GRANDIEN ALF (SE)
FEATURES Location/Qualifiers
source 1. .2867
/organism="Mus sp."
/mol_type="unassigned DNA"
/db_xref="taxon:10095"
ORIGIN

Query Match 100.0%; Score 2867; DB 6; Length 2867;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2867; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         1 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 60
Db         1 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 60
Qy        61 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 120
Db        61 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 120
Qy       121 ACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTT 180
Db       121 ACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTT 180
Qy       181 TATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGAT 240
Db       181 TATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGAT 240
Qy       241 AAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTC 300
Db       241 AAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTC 300
Qy       301 AGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACG 360
Db       301 AGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACG 360
Qy       361 GAGGCGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGA 420
Db       361 GAGGCGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGA 420
Qy       421 CGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGG 480
Db       421 CGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGG 480
Qy       481 AAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACT 540
Db       481 AAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACT 540
```

```
Qy      541 GATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCT 600
Db      541 GATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCT 600

Qy      601 GTATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAA 660
Db      601 GTATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAA 660

Qy      661 GAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAAC 720
Db      661 GAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAAC 720

Qy      721 CGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAAT 780
Db      721 CGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAAT 780

Qy      781 CGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGC 840
Db      781 CGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGC 840

Qy      841 AGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTG 900
Db      841 AGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTG 900

Qy      901 TCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAAT 960
Db      901 TCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAAT 960

Qy      961 ACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCC 1020
Db      961 ACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCC 1020

Qy     1021 AATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGAT 1080
Db     1021 AATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGAT 1080

Qy     1081 CGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTC 1140
Db     1081 CGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTC 1140

Qy     1141 CTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGAT 1200
Db     1141 CTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGAT 1200

Qy     1201 GAGACAAACGGTAGCGCCACCAATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGAT 1260
Db     1201 GAGACAAACGGTAGCGCCACCAATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGAT 1260

Qy     1261 GGCACAGACTGCACAAGCATAGGGACAGTAGAGCAGAAGTCCGGAGAAGACCAGGGCATA 1320
Db     1261 GGCACAGACTGCACAAGCATAGGGACAGTAGAGCAGAAGTCCGGAGAAGACCAGGGCATA 1320

Qy     1321 AAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCAAGAAAAAACTCAAGATATTCCAG 1380
Db     1321 AAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCAAGAAAAAACTCAAGATATTCCAG 1380

Qy     1381 CCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCA 1440
Db     1381 CCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCA 1440

Qy     1441 CAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATG 1500
Db     1441 CAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATG 1500

Qy     1501 AGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAACCCAAGGAAAAGGTCAAGACGAAG 1560
Db     1501 AGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAACCCAAGGAAAAGGTCAAGACGAAG 1560

Qy     1561 CCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTGGGGATTAAAATCTCTTCTGTG 1620
Db     1561 CCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTGGGGATTAAAATCTCTTCTGTG 1620

Qy     1621 CACAAGAGACTAGCGTCAGAGAAAAGGGAAAACCCAGTGAAGAAAGTGATGCTGGCTTCC 1680
Db     1621 CACAAGAGACTAGCGTCAGAGAAAAGGGAAAACCCAGTGAAGAAAGTGATGCTGGCTTCC 1680

Qy     1681 AGGAGTGAGACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGCACACCATCCTGGGCAAGT 1740
Db     1681 AGGAGTGAGACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGCACACCATCCTGGGCAAGT 1740

Qy     1741 GACCACAACTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAGCCCACTGCACTCTCGCCC 1800
Db     1741 GACCACAACTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAGCCCACTGCACTCTCGCCC 1800

Qy     1801 ACCCTATTGAGTAAATGTACGTATCACCCAAAGGCTGGCTTCCCAGGCCCCTCCCATCAT 1860
Db     1801 ACCCTATTGAGTAAATGTACGTATCACCCAAAGGCTGGCTTCCCAGGCCCCTCCCATCAT 1860

Qy     1861 CTGGGTGGCTGCCTGGGGCTGTCTAGGACCAGAGTCCTGGGTGTTCTGGTGCTGATAGTA 1920
Db     1861 CTGGGTGGCTGCCTGGGGCTGTCTAGGACCAGAGTCCTGGGTGTTCTGGTGCTGATAGTA 1920

Qy     1921 GCCAGCAGCTCACTGCCAGCCAGAAGCAGATACCAAGATGCCTCTGGACCCCAGGTGTTC 1980
Db     1921 GCCAGCAGCTCACTGCCAGCCAGAAGCAGATACCAAGATGCCTCTGGACCCCAGGTGTTC 1980

Qy     1981 CTGCCTAGCCTGTGGAGCCTCTCTGGGTGGTTCCTAAAGAGCTGTGTAGGCCTCATGTTG 2040
Db     1981 CTGCCTAGCCTGTGGAGCCTCTCTGGGTGGTTCCTAAAGAGCTGTGTAGGCCTCATGTTG 2040

Qy     2041 GAGGCAATTTCTTATTTCAGTTTTAGGCCTTGGTAAGTATTTGAACTGCTCTCAACAAGA 2100
Db     2041 GAGGCAATTTCTTATTTCAGTTTTAGGCCTTGGTAAGTATTTGAACTGCTCTCAACAAGA 2100

Qy     2101 TGGGACAACATCAGTGCACTGTTCAGACTCAGTTCAGACTTGAGTTCCTCACAGGACAGC 2160
Db     2101 TGGGACAACATCAGTGCACTGTTCAGACTCAGTTCAGACTTGAGTTCCTCACAGGACAGC 2160

Qy     2161 CCAGCACAAACCATGGAGTTTTCCCACGTTACCTATTTTTCTTCAACATGGTCACCACTG 2220
Db     2161 CCAGCACAAACCATGGAGTTTTCCCACGTTACCTATTTTTCTTCAACATGGTCACCACTG 2220

Qy     2221 CTCTTCTATAGAAAAGTGATTTTTTTCATGAGTTAGGGCATTTTTGAAAGGACAGGTAGA 2280
Db     2221 CTCTTCTATAGAAAAGTGATTTTTTTCATGAGTTAGGGCATTTTTGAAAGGACAGGTAGA 2280

Qy     2281 CAATTTGTGTCTATTTCACATACTACACACCTACATAAACAGGTTTGAATTTTGAAGGTC 2340
Db     2281 CAATTTGTGTCTATTTCACATACTACACACCTACATAAACAGGTTTGAATTTTGAAGGTC 2340

Qy     2341 ATCTGCTGCTCAGATCCCATGGTGAGTAATCGTGCTGACCATACACCTACAGCAGTATTC 2400
Db     2341 ATCTGCTGCTCAGATCCCATGGTGAGTAATCGTGCTGACCATACACCTACAGCAGTATTC 2400

Qy     2401 TCATCAAAGTGAGAATTGTGGGTGGTAGCTTCCAGACTCGTAGATGCCCCAGATTTCAAA 2460
Db     2401 TCATCAAAGTGAGAATTGTGGGTGGTAGCTTCCAGACTCGTAGATGCCCCAGATTTCAAA 2460

Qy     2461 TCAGTGTTTATAGTTTGAGTAAACTTTTGAAACCCAAGATTAGATCACTGAGTTTGGAGC 2520
Db     2461 TCAGTGTTTATAGTTTGAGTAAACTTTTGAAACCCAAGATTAGATCACTGAGTTTGGAGC 2520

Qy     2521 CATTCTTGTTTCACTGCTGTACAAGGTACACCTTTCTTACTTTTTGTACAGTGGTTTCAA 2580
Db     2521 CATTCTTGTTTCACTGCTGTACAAGGTACACCTTTCTTACTTTTTGTACAGTGGTTTCAA 2580

Qy     2581 GTGAATTTTCATCTGTGTTCTCATGATAGTGTTTTGTTCCATCTCTGTCCACCCTCAGCC 2640
Db     2581 GTGAATTTTCATCTGTGTTCTCATGATAGTGTTTTGTTCCATCTCTGTCCACCCTCAGCC 2640

Qy     2641 CCAGATAAGGGGGTCTCATGGCCTAAGCTGGCCTTAAATTTTCTGTAAAGTTGGGGGTCC 2700
Db     2641 CCAGATAAGGGGGTCTCATGGCCTAAGCTGGCCTTAAATTTTCTGTAAAGTTGGGGGTCC 2700

Qy     2701 TCCTGCCCCCACCTGTAAAGAGTGCTGAAACTTACAGGCACACACCTGATGTATGCTGTG 2760
```

```
Db       2701 TCCTGCCCCCACCTGTAAAGAGTGCTGAAACTTACAGGCACACACCTGATGTATGCTGTG 2760

Qy       2761 GGGTGTGAACTCAGAGTGTTGGACCAGCAGTCTACCAGCTGAGCTGCAGTTCTAGCCATG 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 GGGTGTGAACTCAGAGTGTTGGACCAGCAGTCTACCAGCTGAGCTGCAGTTCTAGCCATG 2820

Qy       2821 GTAACTGAAACTGCTGATCTTCTTGTCTTTGTTTCCTTTGTGCTGGG 2867
             |||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 GTAACTGAAACTGCTGATCTTCTTGTCTTTGTTTCCTTTGTGCTGGG 2867


RESULT 3
MMU238332
LOCUS       MMU238332               2602 bp    mRNA    linear   ROD 09-JUL-1999
DEFINITION  Mus musculus mRNA for death inducer-obliterator-1 (Dio-1).
ACCESSION   AJ238332
VERSION     AJ238332.1  GI:5457402
KEYWORDS    death inducer-obliterator-1; Dio-1 gene.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Garcia-Domingo,D., Leonardo,E., Grandien,A., Martinez,P.,
            Albar,J.P., Izpisua-Belmonte,J.C. and Martinez-A,C.
  TITLE     DIO-1 is a gene involved in onset of apoptosis in vitro, whose
            misexpression disrupts limb development
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (14), 7992-7997 (1999)
  MEDLINE   99324176
   PUBMED   10393935
REFERENCE   2  (bases 1 to 2602)
  AUTHORS   Garcia-Domingo,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-1999) Garcia-Domingo D., Department of Immunology
            and Oncology, Centro Nacional de Biotecnologia/CSIC, Campus
            Cantoblanco UAM, Ctra. de Colmenar Viejo Km 16, Madrid, 28049,
            SPAIN
FEATURES             Location/Qualifiers
     source          1. .2602
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /cell_line="WOL-1"
                     /cell_type="pre-BI"
     gene            1. .2602
                     /gene="Dio-1"
     CDS             232. .>2073
                     /gene="Dio-1"
                     /function="apoptosis inducer"
                     /codon_start=1
                     /evidence=experimental
                     /product="death inducer-obliterator-1"
                     /protein_id="CAB48401.1"
                     /db_xref="GI:5457403"
                     /db_xref="GOA:Q8C9B9"
                     /db_xref="GOA:Q9WV00"
                     /db_xref="SWISS-PROT:Q8C9B9"
                     /translation="MDDKGHLSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDTE
                     ADPSEQQPQQHNLSLRRSGRQPKRTERVEEFLTTVRRRGKKNVPVSLEDSSEPTSSTV
                     TDVETASEGSVESSSEIRSGPVSDSLGKEHPASSEKAKGGEEEEDTSDSDSDGLTLKE
                     LQNRLRRKREQEPVERSLRGSQNRLRKKRREEDSAETGSVQIGSAEQDRPLCKQEPEA
                     SQGPVSQSETDDIENQLEGKATQGNTEENPREAGKPKPECEVYDPNALYCICRQPHNN
                     RFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETNGSATNEQD
                     SGCRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPSGKKKLKIFQPVVEAPGAP
                     KCIGPGCSSVAQPDSVYCSNDCILKHAAATMRFLSSGKEQKTKPKEKVKTKPEKPSLP
                     KCSVQVGIKISSVHKRLASEKRENPVKKVMLASRSETSGKEAACESSTPSWASDHNYN
                     AVKPEKPEKPTALSPTLLSKCTYHPKAGFPGPSHHLGGCLGLSRTRVLGVLVLIVASS
                     SLPARSRYQDASGPQVFLPSLNSLSGWFLKSCVGLMLEAISYFSFRPW"
ORIGIN

  Query Match           90.8%;  Score 2602;  DB 10;  Length 2602;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 2602;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 60

Qy         61 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 120

Qy        121 ACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTT 180

Qy        181 TATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGAT 240

Qy        241 AAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTC 300

Qy        301 AGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACG 360

Qy        361 GAGGCGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAGGCGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGA 420

Qy        421 CGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGG 480

Qy        481 AAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACT 540

Qy        541 GATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCT 600

Qy        601 GTATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAA 660

Qy        661 GAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAAC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAAC 720

Qy        721 CGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAAT 780

Qy        781 CGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGC 840

Qy        841 AGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTG 900

Qy        901 TCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAAT 960

Qy        961 ACAGAGGAAAACCCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACAGAGGAAAACCCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCC 1020

Qy       1021 AATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGAT 1080
```

```
Qy       1081 CGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTC 1140
Db       1081 CGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTC 1140

Qy       1141 CTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGAT 1200
Db       1141 CTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGAT 1200

Qy       1201 GAGACAAACGGTAGCGCCACCAATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGAT 1260
Db       1201 GAGACAAACGGTAGCGCCACCAATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGAT 1260

Qy       1261 GGCACAGACTGCACAAGCATAGGGACAGTAGAGCAGAAGTCCGGAGAAGACCAGGGCATA 1320
Db       1261 GGCACAGACTGCACAAGCATAGGGACAGTAGAGCAGAAGTCCGGAGAAGACCAGGGCATA 1320

Qy       1321 AAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCAAGAAAAAACTCAAGATATTCCAG 1380
Db       1321 AAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCAAGAAAAAACTCAAGATATTCCAG 1380

Qy       1381 CCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCA 1440
Db       1381 CCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCA 1440

Qy       1441 CAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATG 1500
Db       1441 CAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATG 1500

Qy       1501 AGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAACCCAAGGAAAAGGTCAAGACGAAG 1560
Db       1501 AGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAACCCAAGGAAAAGGTCAAGACGAAG 1560

Qy       1561 CCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTGGGGATTAAAATCTCTTCTGTG 1620
Db       1561 CCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTGGGGATTAAAATCTCTTCTGTG 1620

Qy       1621 CACAAGAGACTAGCGTCAGAGAAAAGGGAAAACCCAGTGAAGAAAGTGATGCTGGCTTCC 1680
Db       1621 CACAAGAGACTAGCGTCAGAGAAAAGGGAAAACCCAGTGAAGAAAGTGATGCTGGCTTCC 1680

Qy       1681 AGGAGTGAGACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGCACACCATCCTGGGCAAGT 1740
Db       1681 AGGAGTGAGACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGCACACCATCCTGGGCAAGT 1740

Qy       1741 GACCACAACTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAGCCCACTGCACTCTCGCCC 1800
Db       1741 GACCACAACTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAGCCCACTGCACTCTCGCCC 1800

Qy       1801 ACCCTATTGAGTAAATGTACGTATCACCCAAAGGCTGGCTTCCCAGGCCCCTCCCATCAT 1860
Db       1801 ACCCTATTGAGTAAATGTACGTATCACCCAAAGGCTGGCTTCCCAGGCCCCTCCCATCAT 1860

Qy       1861 CTGGGTGGCTGCCTGGGGCTGTCTAGGACCAGAGTCCTGGGTGTTCTGGTGCTGATAGTA 1920
Db       1861 CTGGGTGGCTGCCTGGGGCTGTCTAGGACCAGAGTCCTGGGTGTTCTGGTGCTGATAGTA 1920

Qy       1921 GCCAGCAGCTCACTGCCAGCCAGAAGCAGATACCAAGATGCCTCTGGACCCCAGGTGTTC 1980
Db       1921 GCCAGCAGCTCACTGCCAGCCAGAAGCAGATACCAAGATGCCTCTGGACCCCAGGTGTTC 1980

Qy       1981 CTGCCTAGCCTGTGGAGCCTCTCTGGGTGGTTCCTAAAGAGCTGTGTAGGCCTCATGTTG 2040
Db       1981 CTGCCTAGCCTGTGGAGCCTCTCTGGGTGGTTCCTAAAGAGCTGTGTAGGCCTCATGTTG 2040

Qy       2041 GAGGCAATTTCTTATTTCAGTTTTAGGCCTTGGTAAGTATTTGAACTGCTCTCAACAAGA 2100
Db       2041 GAGGCAATTTCTTATTTCAGTTTTAGGCCTTGGTAAGTATTTGAACTGCTCTCAACAAGA 2100

Qy       2101 TGGGACAACATCAGTGCACTGTTCAGACTCAGTTCAGACTTGAGTTCCTCACAGGACAGC 2160
Db       2101 TGGGACAACATCAGTGCACTGTTCAGACTCAGTTCAGACTTGAGTTCCTCACAGGACAGC 2160

Qy       2161 CCAGCACAAACCATGGAGTTTTCCCACGTTACCTATTTTTCTTCAACATGGTCACCACTG 2220
Db       2161 CCAGCACAAACCATGGAGTTTTCCCACGTTACCTATTTTTCTTCAACATGGTCACCACTG 2220

Qy       2221 CTCTTCTATAGAAAAGTGATTTTTTTCATGAGTTAGGGCATTTTTGAAAGGACAGGTAGA 2280
Db       2221 CTCTTCTATAGAAAAGTGATTTTTTTCATGAGTTAGGGCATTTTTGAAAGGACAGGTAGA 2280

Qy       2281 CAATTTGTGTCTATTTCACATACTACACACCTACATAAACAGGTTTGAATTTTGAAGGTC 2340
Db       2281 CAATTTGTGTCTATTTCACATACTACACACCTACATAAACAGGTTTGAATTTTGAAGGTC 2340

Qy       2341 ATCTGCTGCTCAGATCCCATGGTGAGTAATCGTGCTGACCATACACCTACAGCAGTATTC 2400
Db       2341 ATCTGCTGCTCAGATCCCATGGTGAGTAATCGTGCTGACCATACACCTACAGCAGTATTC 2400

Qy       2401 TCATCAAAGTGAGAATTGTGGGTGGTAGCTTCCAGACTCGTAGATGCCCCAGATTTCAAA 2460
Db       2401 TCATCAAAGTGAGAATTGTGGGTGGTAGCTTCCAGACTCGTAGATGCCCCAGATTTCAAA 2460

Qy       2461 TCAGTGTTTATAGTTTGAGTAAACTTTTGAAACCCAAGATTAGATCACTGAGTTTGGAGC 2520
Db       2461 TCAGTGTTTATAGTTTGAGTAAACTTTTGAAACCCAAGATTAGATCACTGAGTTTGGAGC 2520

Qy       2521 CATTCTTGTTTCACTGCTGTACAAGGTACACCTTTCTTACTTTTTGTACAGTGGTTTCAA 2580
Db       2521 CATTCTTGTTTCACTGCTGTACAAGGTACACCTTTCTTACTTTTTGTACAGTGGTTTCAA 2580

Qy       2581 GTGAATTTTCATCTGTGTTCTC 2602
Db       2581 GTGAATTTTCATCTGTGTTCTC 2602
```

```
RESULT 4
AK129117
LOCUS       AK129117                4906 bp    mRNA    linear   ROD 21-NOV-2003
DEFINITION  Mus musculus mRNA for mKIAA0333 protein.
ACCESSION   AK129117
VERSION     AK129117.1  GI:37359897
KEYWORDS    FLI_CDNA.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Okazaki,N., Kikuno,R., Ohara,R., Inamoto,S., Koseki,H., Hiraoka,S.,
            Saga,Y., Nagase,T., Ohara,O. and Koga,H.
  TITLE     Prediction of the coding sequences of mouse homologues of KIAA
            gene: III. the complete nucleotide sequences of 500 mouse
            KIAA-homologous cDNAs identified by screening of terminal sequences
            of cDNA clones randomly sampled from size-fractionated libraries
  JOURNAL   DNA Res. 10 (4), 167-180 (2003)
  MEDLINE   22977043
   PUBMED   14621295
REFERENCE   2  (bases 1 to 4906)
  AUTHORS   Okazaki,N., Kikuno,R., Nagase,T., Ohara,O. and Koga,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-2003) Hisashi Koga, Kazusa DNA Research
            Institute, Laboratory for Genome Informatics; 2-6-7
            Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan
            (E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)
COMMENT     The CREATE program supported by Japan science and technology
            corporation; cDNA full insert sequencing: Kazusa DNA Research
            Institute; cDNA library construction, clone selection and 5'- &
            3'-end one pass sequencing.
FEATURES             Location/Qualifiers
     source          1. .4906
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="mbg16015"
                     /tissue_type="brain"
                     /dev_stage="adult"
                     /note="vector:modified pBC SK+"
```

```
     gene            1. .4906
                     /gene="mKIAA0333"
     CDS             <224. .3829
                     /gene="mKIAA0333"
                     /note="CDS is predicted by in silico analysis. Start codon
                     is not identified."
                     /codon_start=1
                     /evidence=not_experimental
                     /product="mKIAA0333 protein"
                     /protein_id="BAC97927.1"
                     /db_xref="GI:37359898"
                     /translation="FYVGFPAKGFQAFSFGTGMDDKGHLSNEEAPKAIKPTSKEFRKT
                     WGFRRTTIAKREGAGDTEVDPSEQQPQQHNLSLRRSGRQPKRTERVEEFLTTVRRRGK
                     KNVPVSLEDSSEPTSSTVTDVETASEGSVESSSSEIRSGPVSDSLGKEHPASSEKAKGG
                     EEEEDTSDSDSDGLTLKELQNRLRRKREQEPVERSLRGSQNRLRKKRREEDSAETGSV
                     QIGSAEQDRPLCKQEPEASQGPVSQSETDDIENQLEGKATQGNTEENPREAGKPKPEC
                     EVYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNC
                     TILQVQDETNGSATDEQDSGCRSVGADGTDCTSIGTVEQKSGEDQGIKGRIEKAANPS
                     GKKKLKIFQPVVEAPGAPKCIGPGCSSVAQPDSVYCSNDCILKHAAATMRFLSSGKEQ
                     KTKPKEKVKTKPEKFSLPKCSVQVGIKISSVHKRLASEKRENPVKKVMLASRSETSGK
                     EAACESSTPSWASDHNYNAVKPEKPEKPTALSPTLLSKSMKDDRRVEDRTMAAVTIPK
                     KALPSASLVGRQTSPRNLVPKKLPPYSNMAGAKPAIKKLPSGFKGTIPKRPWPSATLS
                     GTSARQAGPTPMTAASKKLPGSAAVVGVTRKPMSANVPAASPAPGRLGPVSPAPSQPN
                     SQIRQNIRRSLKBILWKRVNDSDDLIMTENEVGKIALHIEKEMFNLFQVTDNRYKSKY
                     RSIMFNLKDPKNQGLFHRVLREEISLAKLVRMKPEELVSKELSMWTEKPTKSVIESRT
                     KLLNESKKNTTKPETIPDMEDSPPVSDSEEQQESVRAAPEKSAAPLLDVFSSMLKDTT
                     SQHRAHLFDLNCKICTGQVPSSEDEPAPKKQKLSASSKKEDFKPRHDSSPPNAVPNTA
                     DEGIADTLPENASEPDPESTSSLNQERKCFPESPGDSHPEPSSLGGLSPSSASGGSGV
                     VTTVTMSGRDPRTALSGSCTVTASMAAHLDNSQASETKLDMIKPALTSAVVPKSILAK
                     PSSSPDPRYLSVPPSPSISESRSPPEGDTTLFLSRLNTIWKGFINMQSVAKFVTKAYP
                     VSGCLDYLSEDLPDTIHIGGRIAPKTVWDYVGKLKSSVSKELCLIRFHPATEEEEVAY
                     ISLYSYFSSRGRFGVVANNNRHVKDLYLIPLSAKDPVPSKLLPFEGPGKHPVSGR"
ORIGIN

  Query Match           61.8%;  Score 1770.4;  DB 10;  Length 4906;
  Best Local Similarity  99.7%;  Pred. No. 0;
  Matches 1774;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy         44 GGAGCTTGTTTAAGAGGCAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTT 103
Db         90 GGAGCTTGTTTAAGAGGCAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTT 149

Qy        104 CTCATTGATGAACTGTGACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGG 163
Db        150 CTCATTGATGAACTGTGACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGG 209

Qy        164 CCCCAGGTCTATAATTTTATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTG 223
Db        210 CCCCAGGTCTATAATTTTATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTG 269

Qy        224 GGACAGGTATGGATGATAAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAAC 283
Db        270 GGACAGGTATGGATGATAAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAAC 329

Qy        284 CCACCAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTG 343
Db        330 CCACCAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTG 389

Qy        344 AGGGTGCAGGAGACACGGAGGCGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCT 403
Db        390 AGGGTGCAGGAGACACGGAGGTGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCT 449

Qy        404 CCCTGCGCCGCAGTGGACGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCA 463
Db        450 CCCTGCGCCGCAGTGGACGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCA 509

Qy        464 CGGTTCGGCGCCGAGGGAAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCA 523
Db        510 CGGTTCGGCGCCGAGGGAAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCA 569

Qy        524 CATCTTCCACAGTCACTGATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTG 583
Db        570 CATCTTCCACAGTCACTGATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTG 629

Qy        584 AGATCAGAAGTGGCCCTGTATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAAA 643
Db        630 AGATCAGAAGTGGCCCTGTATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAAA 689

Qy        644 AGGCAAAAGGAGGTGAAGAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACGT 703
Db        690 AGGCAAAAGGAGGTGAAGAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACGT 749

Qy        704 TGAAGGAACTTCAGAACCGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCCC 763
Db        750 TGAAGGAACTTCAGAACCGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCCC 809

Qy        764 TGAGAGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTG 823
Db        810 TGAGAGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTG 869

Qy        824 GGAGTGTCCAAATAGGCAGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGG 883
Db        870 GGAGTGTCCAAATAGGCAGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGG 929

Qy        884 CTAGTCAGGGACCAGTGTCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGA 943
Db        930 CTAGTCAGGGACCAGTGTCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGA 989

Qy        944 AGGCGACTCAGGGAAATACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAGT 1003
Db        990 AGGCGACTCAGGGAAATACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAGT 1049

Qy       1004 GTGAGGTTTACGACCCCAATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGT 1063
Db       1050 GTGAGGTTTACGACCCCAATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGT 1109

Qy       1064 TTATGATCTGCTGTGATCGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCTG 1123
Db       1110 TTATGATCTGCTGTGATCGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCTG 1169

Qy       1124 AGGCCCGAGGGCGGCTCCTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTGCACCA 1183
Db       1170 AGGCCCGAGGGCGGCTCCTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATTGCACCA 1229

Qy       1184 TTTTGCAAGTGCAGGATGAGACAAACGGTAGCGCCACCAATGAGCAGGACTCTGGGTGCA 1243
Db       1230 TTTTGCAAGTGCAGGATGAGACAAACGGTAGCGCCACCGATGAGCAGGACTCTGGGTGCA 1289

Qy       1244 GATCTGTGGGTGCTGATGGCACAGACTGCACAAGCATAGGGACAGTAGAGCAGAAGTCCG 1303
Db       1290 GATCTGTGGGTGCTGATGGCACAGACTGCACAAGCATAGGGACAGTAGAGCAGAAGTCCG 1349

Qy       1304 GAGAAGACCAGGGCATAAAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCAAGAAAA 1363
Db       1350 GAGAAGACCAGGGCATAAAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCAAGAAAA 1409

Qy       1364 AACTCAAGATATTCCAGCCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTGGCCCTG 1423
Db       1410 AACTCAAGATATTCCAGCCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTGGCCCTG 1469

Qy       1424 GGTGTTCCAGTGTAGCACAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTCTCAAAC 1483
Db       1470 GGTGTTCCAGTGTAGCACAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTCTCAAAC 1529

Qy       1484 ACGCAGCAGCTACCATGAGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAACCCAAGG 1543
Db       1530 ACGCAGCAGCTACCATGAGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAACCCAAGG 1589

Qy       1544 AAAAGGTCAAGACGAAGCCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTGGGGA 1603
Db       1590 AAAAGGTCAAGACGAAGCCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTGGGGA 1649

Qy       1604 TTAAAATCTCTTCTGTGCACAAGAGACTAGCGTCAGAGAAAAGGGAAAACCCAGTGAAGA 1663
Db       1650 TTAAAATCTCTTCTGTGCACAAGAGACTAGCGTCAGAGAAAAGGGAAAACCCAGTGAAGA 1709
```

```
Qy       1664 AAGTGATGCTGGCTTCCAGGAGTGAGACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGCA 1723
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1710 AAGTGATGCTGGCTTCCAGGAGTGAGACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGCA 1769

Qy       1724 CACCATCCTGGGCAAGTGACCACAACTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAGC 1783
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1770 CACCATCCTGGGCAAGTGACCACAACTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAGC 1829

Qy       1784 CCACTGCACTCTCGCCCACCCTATTGAGTAAATGTACGTA 1823
             |||||||||||||||||||||||||||||||||  | | |
Db       1830 CCACTGCACTCTCGCCCACCCTATTGAGTAAATCCATGAA 1869


RESULT 5
AL732560/c
LOCUS       AL732560              104663 bp    DNA     linear   ROD 22-AUG-2002
DEFINITION  Mouse DNA sequence from clone RP24-209P15 on chromosome 2, complete
            sequence.
ACCESSION   AL732560
VERSION     AL732560.11  GI:22474418
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 104663)
  AUTHORS   Leongamornlert,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 23, 2002 this sequence version replaced gi:22204532.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP24-209P15 is
            from a Male (C57BL/6J) mouse BAC LIbrary VECTOR: pTARBAC1.
FEATURES             Location/Qualifiers
     source          1. .104663
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /clone="RP24-209P15"
                     /clone_lib="RPCI-24"
ORIGIN

  Query Match             43.3%;  Score 1241.2;  DB 10;  Length 104663;
  Best Local Similarity   99.1%;  Pred. No. 0;
  Matches 1269;  Conservative    0;  Mismatches    8;  Indels    3;  Gaps    2;

Qy       1588 AGTGTTCAGGTGGGGATTAAAATCTCTTCTGTGCACAAGAGACTAGCGTCAGAGAAAAGG 1647
              | | | ||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      46336 ATTATCCAGGTGGGGATTAAAATCTCTTCTGTGCACAAGAGACTAGCGTCAGAGAAAAGG 46277

Qy       1648 GAAAACCCAGTGAAGAAAGTGATGCTGGCTTCCAGGAGTGAGACTTCTGGGAAGGAGGCA 1707
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      46276 GAAAACCCAGTGAAGAAAGTGATGCTGGCTTCCAGGAGTGAGACTTCTGGGAAGGAGGCA 46217

Qy       1708 GCCTGTGAGAGCAGCACACCATCCTGGGCAAGTGACCACAACTACAATGCTGTGAAGCCA 1767
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      46216 GCCTGTGAGAGCAGCACACCATCCTGGGCAAGTGACCACAACTACAATGCTGTGAAGCCA 46157

Qy       1768 GAGAAGCCAGAGAAGCCCACTGCACTCTCGCCCACCCTATTGAGTAAATGTACGTATCAC 1827
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      46156 GAGAAGCCAGAGAAGCCCACTGCACTCTCGCCCACCCTATTGAGTAAATGTACGTATCAC 46097

Qy       1828 CCAAAGGCTGGCTTCCCAGGCCCCTCCCATCATCTGGGTGGCTGCCTGGGGCTGTCTAGG 1887
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      46096 CCAAAGGCTGGCTTCCCAGGCCCCTCCCATCATCTGGGTGGCTGCCTGGGGCTGTCTAGG 46037

Qy       1888 ACCAGAGTCCTGGGTGTTCTGGTGCTGATAGTAGCCAGCAGCTCACTGCCAGCCAGAAGC 1947
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      46036 ACCAGAGTCCTGGGTGTTCTGGTGCTGATAGTAGCCAGCAGCTCACTGCCAGCCAGAAGC 45977

Qy       1948 AGATACCAAGATGCCTCTGGACCCCAGGTGTTCCTGCCTAGCCTGTGGAGCCTCTCTGGG 2007
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45976 AGATACCAAGATGCCTCTGGACCCCAGGTGTTCCTGCCTAGCCTGTGGAGCCTCTCTGGG 45917

Qy       2008 TGGTTCCTAAAGAGCTGTGTAGGCCTCATGTTGGAGGCAATTTCTTATTTCAGTTTTAGG 2067
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45916 TGGTTCCTAAAGAGCTGTGTAGGCCTCATGTTGGAGGCAATTTCTTATTTCAGTTTTAGG 45857

Qy       2068 CCTTGGTAAGTATTTGAACTGCTCTCAACAAGATGGGACAACATCAGTGCACTGTTCAGA 2127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45856 CCTTGGTAAGTATTTGAACTGCTCTCAACAAGATGGGACAACATCAGTGCACTGTTCAGA 45797

Qy       2128 CTCAGTTCAGACTTGAGTTCCTCACAGGACAGCCCAGCACAAACCATGGAGTTTTCCCAC 2187
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      45796 CTCAGTTCAGACTTGAGTTCCTCACAGGACAGCCCAGCACAAACCAT-GAGTTTTCCCAC 45738

Qy       2188 GTTACCTATTTTTCTTCAACATGGTCACCACTGCTCTTCTATAGAAAAGTGATTTTTTTC 2247
              ||||||||||||   | ||||||||||||||||||||||||||||||||||||||||||||
Db      45737 GTTACCTATTTT--CTTCACATGGTCACCACTGCTCTTCTATAGAAAAGTGATTTTTTTC 45680

Qy       2248 ATGAGTTAGGGCATTTTTGAAAGGACAGGTAGACAATTTGTGTCTATTTCACATACTACA 2307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45679 ATGAGTTAGGGCATTTTTGAAAGGACAGGTAGACAATTTGTGTCTATTTCACATACTACA 45620

Qy       2308 CACCTACATAAACAGGTTTGAATTTTGAAGGTCATCTGCTGCTCAGATCCCATGGTGAGT 2367
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45619 CACCTACATAAACAGGTTTGAATTTTGAAGGTCATCTGCTGCTCAGATCCCATGGTGAGT 45560

Qy       2368 AATCGTGCTGACCATACACCTACAGCAGTATTCTCATCAAAGTGAGAATTGTGGGTGGTA 2427
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45559 AATCGTGCTGACCATACACCTACAGCAGTATTCTCATCAAAGTGAGAATTGTGGGTGGTA 45500

Qy       2428 GCTTCCAGACTCGTAGATGCCCCAGATTTCAAATCAGTGTTTATAGTTTGAGTAAACTTT 2487
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45499 GCTTCCAGACTCGTAGATGCCCCAGATTTCAAATCAGTGTTTATAGTTTGAGTAAACTTT 45440

Qy       2488 TGAAACCCAAGATTAGATCACTGAGTTTGGAGCCATTCTTGTTTCACTGCTGTACAAGGT 2547
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45439 TGAAACCCAAGATTAGATCACTGAGTTTGGAGCCATTCTTGTTTCACTGCTGTACAAGGT 45380

Qy       2548 ACACCTTTCTTACTTTTTGTACAGTGGTTTCAAGTGAATTTTCATCTGTGTTCTCATGAT 2607
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45379 ACACCTTTCTTACTTTTTGTACAGTGGTTTCAAGTGAATTTTCATCTGTGTTCTCATGAT 45320

Qy       2608 AGTGTTTTGTTCCATCTCTGTCCACCCTCAGCCCCAGATAAGGGGGTCTCATGGCCTAAG 2667
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db      45319 AGTGTTTTGTTCCATCTCTGTCCACCCTCAGCCCCAGATAAGGGGGTCTCATGGCCTAGG 45260

Qy       2668 CTGGCCTTAAATTTTCTGTAAAGTTGGGGGTCCTCCTGCCCCCACCTGTAAAGAGTGCTG 2727
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45259 CTGGCCTTAAATTTTCTGTAAAGTTGGGGGTCCTCCTGCCCCCACCTGTAAAGAGTGCTG 45200
```

```
Qy       2728 AAACTTACAGGCACACACCTGATGTATGCTGTGGGGTGTGAACTCAGAGTGTTGGACCAG 2787
             ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db      45199 AAACTTACAGGCACACACCTGATGTATGTTGTGGGGTGTGAACTCAGAGTGTTGGACCAG 45140

Qy       2788 CAGTCTACCAGCTGAGCTGCAGTTCTAGCCATGGTAACTGAAACTGCTGATCTTCTTGTC 2847
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45139 CAGTCTACCAGCTGAGCTGCAGTTCTAGCCATGGTAACTGAAACTGCTGATCTTCTTGTC 45080

Qy       2848 TTTGTTTCCTTTGTGCTGGG 2867
             ||||||||||||||||||||
Db      45079 TTTGTTTCCTTTGTGCTGGG 45060
```

RESULT 6
BD244712
LOCUS BD244712 2610 bp DNA linear PAT 17-JUL-2003
DEFINITION Genes encoding for the human and murine death inducer-obliterator-1.
ACCESSION BD244712
VERSION BD244712.1 GI:33054482
KEYWORDS JP 2002526040-A/1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2610)
AUTHORS Alonso,C.M., Domingo,D.G., Grandien,A., Leonardo,E. and Martinez,P.
TITLE Genes encoding for the human and murine death inducer-obliterator-1
JOURNAL Patent: JP 2002526040-A 1 20-AUG-2002;
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS
COMMENT OS Homo sapiens (human)
PN JP 2002526040-A/1
PD 20-AUG-2002
PF 10-SEP-1999 JP 2000570314
PR 10-SEP-1998 SE 9803069-5,17-SEP-1998 US 60/100873 PI
CARLOS MARTINEZ ALONSO,DAVID GARCIA DOMINGO,ALF GRANDIEN, PI
ESTHER LEONARDO,
PI PEDRO MARTINEZ
PC C12N15/09,A61K38/00,A61K45/00,A61K48/00,A61P3/10,A61P17/00, PC A61P29/00,
PC A61P35/00,A61P37/06,A61P43/00,C07K14/47,C07K16/18,C12N1/21, PC C12N5/10,
PC C12P21/02,C12Q1/02,C12Q1/68,G01N33/15,G01N33/50//C12P21/08, PC C12N15/00,
PC C12N5/00,A61K37/02
CC This gene is referred to in the application as Human Death Inducer CC
CC Obliterator Gene 1.
CC It has now been named by the Human Gene
Nomenclature Committee
CC as
CC Death-Associated Transcription Factor (DATF-1) FH Key Location/Qualifiers
FT source 1. .2610
FT /organism='Homo sapiens (human)'.
FEATURES Location/Qualifiers
source 1. .2610
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN

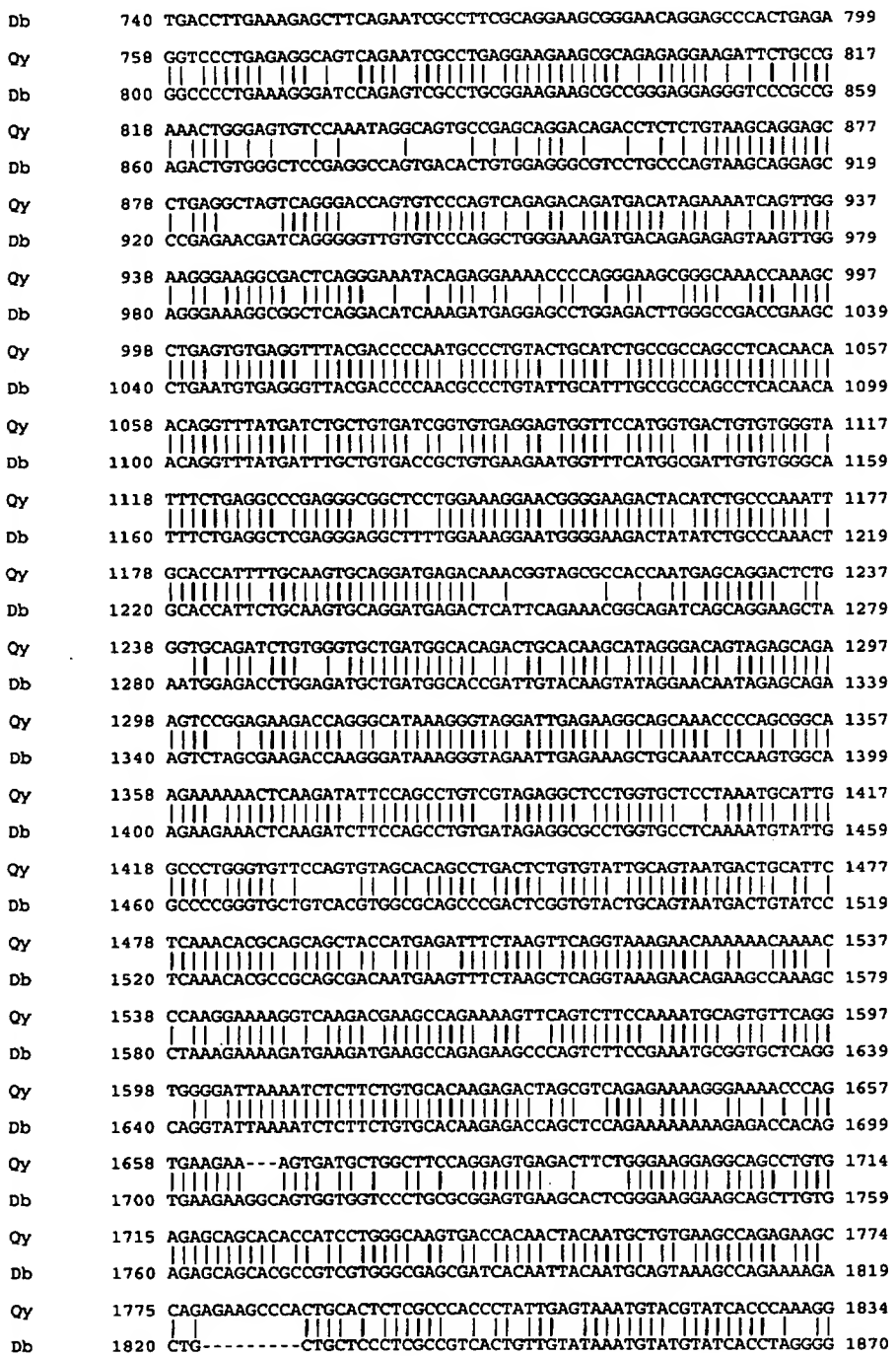
Query Match 35.7%; Score 1023.6; DB 6; Length 2610;
Best Local Similarity 71.9%; Pred. No. 3.7e-303;
Matches 1494; Conservative 0; Mismatches 534; Indels 50; Gaps 10;

```
Qy          1 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 60
              |||| ||      |||||||||||||| || ||  | |||| | |   |   |  ||
Db         21 TCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGCCTGCGGAGCTTACTCCACGG 80

Qy         61 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 120
                |   || ||      |  |   |    |      ||| | | | | |  ||||| |||
```

```
Db         81 GAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTGTTACTCGTGAACAGTG 140

Qy        121 ACTGACAACAGTGGGGTGAG--GCTTGGCCGTCTGCTTGACC----------TGGCCC 166
              |||||||||||  ||    || |||| ||||||||  |          |   ||
Db        141 GCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCCAGAGGTTTCGTCTGCC 200

Qy        167 CAGGTCTATAATTTTATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGA 226
                |||  |  || ||| |||||||| |||  ||| || |||||||||||||| ||||
Db        201 AGGGTTTTTGGTTGTATTTAGGATTT-CAGGGAAAAGTGTCCAAGCTTTCAGTGTTGGAG 259

Qy        227 CAGGTATGGATGATAAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCA 286
              |||||||||| || |||||  ||| |||||||||||||  |||||  |||||  |||||||||||
Db        260 CAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTAAGGCCATCAAACCCA 319

Qy        287 CCAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGG 346
              |||| || |||||||||||||| |||||||||||||| |||||  ||  |||||  || ||||
Db        320 CCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTATCGCCAAGCGAGAGG 379

Qy        347 GTGCAGGAGACACGGAGGCGGACCCCAGTGAGCAGCAACC------ACAGCAGCATAACC 400
              | |||||  ||| |||||||  |||||     ||||  ||  |||        |||||||     ||
Db        380 GCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCACAGCAGCAGCTGGGCC 439

Qy        401 TCTCCCTGCGCCGCAGTGGACGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTA 460
              | |||||||| |||||||  ||||  ||  ||  ||  ||||||  |  ||  ||  ||||  || |
Db        440 TGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGA 499

Qy        461 CCACGGTTCGGCGCCGAGGGAAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGC 520
              |||  |  ||||||||  ||  |  ||   ||||  ||  ||||||||||||||  ||||||
Db        500 CCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGC 559

Qy        521 CCACATCTTCCACAGTCACTGATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTT 580
              ||||  ||  |  |  |  |||  ||  |  ||||||||  ||||||||  |||||  ||||||  ||
Db        560 CCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCAGCGTGGAAAGCGCTT 619

Qy        581 CTGAGATCAGAAGTGGCCCTGTATCTGACTCCTTA---GGGAAAGAACATCCTGCCTCTT 637
              ||||||  ||||||  |||||     ||||  |||  |  |||  ||||  ||  ||||||||
Db        620 CTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGGAACGACCAGCCTCTT 679

Qy        638 CTGAAAAGGCAAAAGGAGGTGAAGAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCC 697
              |||||||||  ||||||||  ||  ||  |  ||  ||||||||  ||  ||||||||  |||||||
Db        680 CTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGGCC 739

Qy        698 TTACGTTGAAGGAACTTCAGAACCGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGA 757
              | ||  |||||  ||  ||||||||  ||||||||  ||  |||||  ||  ||  ||  ||||
Db        740 TGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAGGAAGCGGGAACAGGAGCCCACTGAGA 799

Qy        758 GGTCCCTGAGAGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCG 817
              ||  ||||||  |||  |  ||||  |||||||  ||||||||||  |  |||||  |  |  ||||
Db        800 GGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAAGAAGCGCCGGGAGGAGGGTCCCGCCG 859

Qy        818 AAACTGGGAGTGTCCAAATAGGCAGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGC 877
              |  ||||  |  |  |  |  |  |  |  |||  |  |  |  |  ||||||||||||||
Db        860 AGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCCTGCCCAGTAAGCAGGAGC 919

Qy        878 CTGAGGCTAGTCAGGGACCAGTGTCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGG 937
              |  |||      ||||||     |||||||||  |  |  ||  ||||||||  |||  |  ||||||
Db        920 CCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGGGAAAGATGACAGAGAGAGTAAGTTGG 979

Qy        938 AAGGGAAGGCGACTCAGGGAAATACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGC 997
              |  ||  ||||||  ||||||  |  |  |||  ||  |  ||  |  ||  ||||  |||  ||||
Db        980 AGGGAAAGGCGGCTCAGGACATCAAAGATGAGGAGCCTGGAGACTTGGGCCGACCGAAGC 1039

Qy        998 CTGAGTGTGAGGTTTACGACCCCAATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACA 1057
              ||||  |||||||  ||||||||||  |||||  ||||||||||||||||||||||||||
Db       1040 CTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAACA 1099

Qy       1058 ACAGGTTTATGATCTGCTGTGATCGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTA 1117
              ||||||||||||||  ||||||||  ||  |||||  ||  |||||  |||||  ||  ||||||||  |
Db       1100 ACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGCA 1159

Qy       1118 TTTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAACGGGGAAGACTACATCTGCCCAAATT 1177
              ||||||||||  ||||||  ||||  ||||||||||  ||||||||||  ||||||||||||  |
Db       1160 TTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAACT 1219
```

```
Qy       1178 GCACCATTTTGCAAGTGCAGGATGAGACAAACGGTAGCGCCACCAATGAGCAGGACTCTG 1237
Db       1220 GCACCATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTA 1279

Qy       1238 GGTGCAGATCTGTGGGTGCTGATGGCACAGACTGCACAAGCATAGGGACAGTAGAGCAGA 1297
Db       1280 AATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAGGAACAATAGAGCAGA 1339

Qy       1298 AGTCCGGAGAAGACCAGGGCATAAAGGGTAGGATTGAGAAGGCAGCAAACCCCAGCGGCA 1357
Db       1340 AGTCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTGCAAATCCAAGTGGCA 1399

Qy       1358 AGAAAAAACTCAAGATATTCCAGCCTGTCGTAGAGGCTCCTGGTGCTCCTAAATGCATTG 1417
Db       1400 AGAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGGCGCCTGGTGCCTCAAAATGTATTG 1459

Qy       1418 GCCCTGGGTGTTCCAGTGTAGCACAGCCTGACTCTGTGTATTGCAGTAATGACTGCATTC 1477
Db       1460 GCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTGCAGTAATGACTGTATCC 1519

Qy       1478 TCAAACACGCAGCAGCTACCATGAGATTTCTAAGTTCAGGTAAAGAACAAAAAACAAAAC 1537
Db       1520 TCAAACACGCCGCAGCGACAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAAGCCAAAGC 1579

Qy       1538 CCAAGGAAAAGGTCAAGACGAAGCCAGAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGG 1597
Db       1580 CTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCCCAGTCTTCCGAAATGCGGTGCTCAGG 1639

Qy       1598 TGGGGATTAAAATCTCTTCTGTGCACAAGAGACTAGCGTCAGAGAAAAGGGAAAACCCAG 1657
Db       1640 CAGGTATTAAAATCTCTTCTGTGCACAAGAGACCAGCTCCAGAAAAAAAAGAGACCACAG 1699

Qy       1658 TGAAGAA---AGTGATGCTGGCTTCCAGGAGTGAGACTTCTGGGAAGGAGGCAGCCTGTG 1714
Db       1700 TGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTG 1759

Qy       1715 AGAGCAGCACACCATCCTGGGCAAGTGACCACAACTACAATGCTGTGAAGCCAGAGAAGC 1774
Db       1760 AGAGCAGCACGCCGTCGTGGGCGAGCGATCACAATTACAATGCAGTAAAGCCAGAAAAGA 1819

Qy       1775 CAGAGAAGCCCACTGCACTCTCGCCCACCCTATTGAGTAAATGTACGTATCACCCAAAGG 1834
Db       1820 CTG---------CTGCTCCCTCGCCGTCACTGTTGTATAAATGTATGTATCACCTAGGGG 1870

Qy       1835 CTGGCTTCCCAGGCCCCTCCCATCATCT-----GGGTGGCTGCCTGGGGCTGTCTAGGAC 1889
Db       1871 TTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGATAGCCATCCCCTGGGCCTGTCCAGGAC 1930

Qy       1890 CAG-----AGTCCTGGGTGTTCTGGTGCTGATAGTAGCCAGCAGC----TCACTGCCAGC 1940
Db       1931 TGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAGACACCGGCTGCACCATCAGCGGGAAG 1990

Qy       1941 CAGAAGCAGATACCAAGATGCCTCTGGACCCCAGGTGTTCCTGCCTAGCCTGTGGAGCCT 2000
Db       1991 CAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCTGTGTCCATCCCTAGTCTGTCAGGACT 2050

Qy       2001 CTCTGGGTGGTTCCTAAAGAGCTGTGTAGGCCTCATGT 2038
Db       2051 TCCTGTCACTGTTTTCCAAAGCTGTAAACCTCACTGGT 2088
```

```
RESULT 7
AX023369
LOCUS       AX023369               2610 bp    DNA     linear   PAT 15-SEP-2000
DEFINITION  Sequence 1 from Patent WO0015787.
ACCESSION   AX023369
VERSION     AX023369.1  GI:10183781
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Leonardo,E., Martinez,P., Alonso,C.M., Domingo,D.G. and Grandien,A.
  TITLE     Genes encoding for the human and murine death in
            ducer-obliterator-1
  JOURNAL   Patent: WO 0015787-A 1 23-MAR-2000;
            LEONARDO ESTHER (ES) ; MARTINEZ PEDRO (ES) ; CONSEJO SUPERIOR
            INVESTIGACION (ES) ; DOMINGO DAVID GARCIA (ES) ; MARTINEZ ALONSO
            CARLOS (ES) ; BANNERMAN DAVID GARDNER (GB) ; GRANDIEN ALF (SE)
FEATURES             Location/Qualifiers
     source          1. .2610
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
                     /note="This gene is referred to in the application as
                     Human Death Inducer Obliterator Gene 1. It has now been
                     named by the Human Gene Nomenclature Committee as
                     Death-Associated Transcription Factor (DATF-1)"
ORIGIN

  Query Match           35.7%;  Score 1023.6;  DB 6;  Length 2610;
  Best Local Similarity  71.9%;  Pred. No. 3.7e-303;
  Matches 1494;  Conservative    0;  Mismatches  534;  Indels   50;  Gaps   10;

Qy          1 TCCGTGGTAGCTCTGGAAATGGCTGCGGGATCCCGGCGGCCGGGGAGCTTGTTTAAGAGG 60
Db         21 TCCGCGGCGTTCGGGGAAATGGCTGCGAGACCCTAGAGGCCTGCGGAGCTTACTCCACGG 80

Qy         61 CAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGATGAACTGTG 120
Db         81 GAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAAGCCTGTTACTCGTGAACAGTG 140

Qy        121 ACTGACAACAGTGGGGTGAG--GCTTGGCCGTCTGCTTGACC------------TGGCCC 166
Db        141 GCTGACAACAGTGTTGTTGTGAGCCTGGCTGTCTGCTTGGACCCAGAGGTTTCGTCTGCC 200

Qy        167 CAGGTCTATAATTTTATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGA 226
Db        201 AGGGTTTTTGGTTGTATTTAGGATTT-CAGGGAAAAGTGTCCAAGCTTTCAGTGTTGGAG 259

Qy        227 CAGGTATGGATGATAAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCA 286
Db        260 CAGGTATGGACGACAAAGGCGACCCGAGCAATGAGGAGGCACCTAAGGCCATCAAACCCA 319

Qy        287 CCAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGG 346
Db        320 CCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCGAAGGACCACTATCGCCAAGCGAGAGG 379

Qy        347 GTGCAGGAGACACGGAGGCGGACCCCAGTGAGCAGCAACC------ACAGCAGCATAACC 400
Db        380 GCGCAGGGGACGCGGAGGCTGACCCACTGGAGCCGCCACCCCCACAGCAGCAGCTGGGCC 439

Qy        401 TCTCCCTGCGCCGCAGTGGACGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTA 460
Db        440 TGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGA 499

Qy        461 CCACGGTTCGGCGCCGAGGGAAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGC 520
Db        500 CCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCCTGTCTCCCTGGAGGATTCTGGTGAGC 559

Qy        521 CCACATCTTCCACAGTCACTGATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTT 580
Db        560 CCACGTCCTGCCCCGCCACAGACGCCGAGACAGCCTCCGAGGGCAGCGTGGAAAGCGCTT 619

Qy        581 CTGAGATCAGAAGTGGCCCTGTATCTGACTCCTTA---GGGAAAGAACATCCTGCCTCTT 637
Db        620 CTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCACAGCTGTGAAGGAACGACCAGCCTCTT 679

Qy        638 CTGAAAAGGCAAAAGGAGGTGAAGAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCC 697
Db        680 CTGAAAAGGTGAAAGGAGGGGATGACCACGATGACACCTCCGATAGTGACAGCGATGGCC 739

Qy        698 TTACGTTGAAGGAACTTCAGAACCGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGA 757
```



RESULT 8
BC004237
LOCUS BC004237 2767 bp mRNA linear PRI 12-NOV-2003
DEFINITION Homo sapiens death associated transcription factor 1, transcript variant 1, mRNA (cDNA clone MGC:896 IMAGE:2959493), complete cds.
ACCESSION BC004237
VERSION BC004237.2 GI:33872756
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2767)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE 22388257
PUBMED 12477932
REFERENCE 2 (bases 1 to 2767)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT On Aug 19, 2003 this sequence version replaced gi:13278980.
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbiology.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

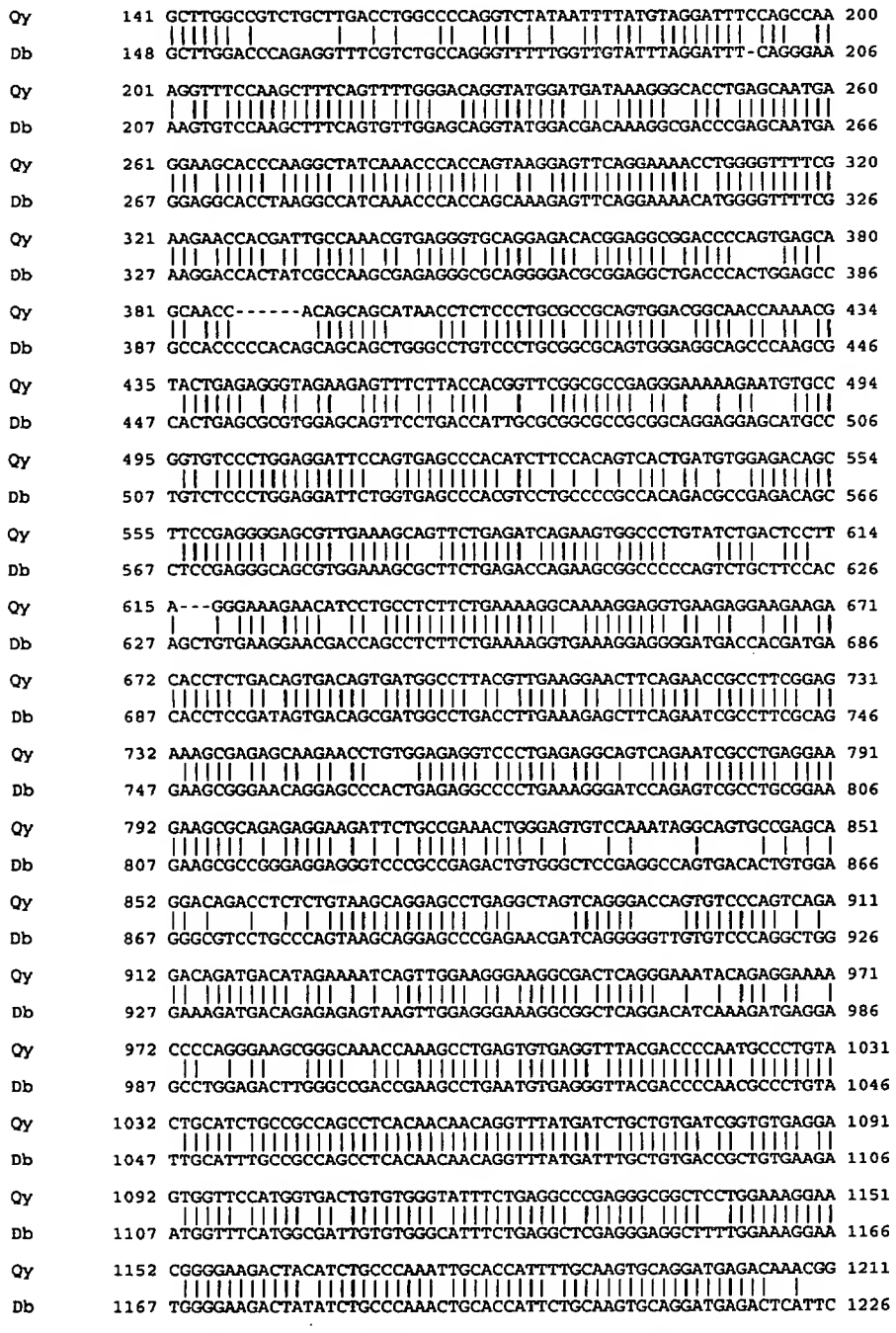
```
          Clone distribution: MGC clone distribution information can be found
          through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
          Series: IRAL Plate: 2 Row: o Column: 16
          This clone was selected for full length sequencing because it
          passed the following selection criteria: matched mRNA gi: 18375620.
FEATURES             Location/Qualifiers
     source          1. .2767
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:896 IMAGE:2959493"
                     /tissue_type="Colon, adenocarcinoma"
                     /clone_lib="NIH_MGC_15"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     gene            1. .2767
                     /gene="DATF1"
                     /note="synonyms: DIO1, FLJ11265, KIAA0333"
                     /db_xref="LocusID:11083"
                     /db_xref="MIM:604140"
     CDS             282. .1970
                     /codon_start=1
                     /product="death associated transcription factor 1, isoform
                     a"
                     /protein_id="AAH04237.1"
                     /db_xref="GI:13278981"
                     /db_xref="LocusID:11083"
                     /translation="MDDKGDPSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDAE
                     ADPLEPPPPQQQLGLSLRRSGRQPKRTERVEQFLTIARRRGRRSMPVSLEDSGEPTSC
                     PATDAETASEGSVESASETRSGPQSASTAVKERPASSEKVKGGDDHDDTSDSDSDGLT
                     LKELQNRLRRKREQEPTERPLKGIQSRLRKKRREEGPAETVGSEASDTVEGVLPSKQE
                     PENDQGVVSQAGKDDRESKLEGKAAQDIKDEEPGDLGRPKPECEGYDPNALYCICRQP
                     HNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETHSETAD
                     QQEAKWRPGDADGTDCTSIGTIEQKSSEDQGIKGRIEKAANPSGKKKLKIFQPVIEAP
                     GASKCIGPGCCHVAQPDSVYCSNDCILKHAAATMKFLSSGKEQKPKPKEKMKMKPEKP
                     SLPKCGAQAGIKISSVHKRPAPEKKETTVKKAVVVPARSEALGKEAACESSTPSWASD
                     HNYNAVKPEKTAAPSPSLLYKCMYHLGVGLLDPSRSFWIAIPWACPGLGVAALC"
     misc_feature    1092. .1241
                     /note="PHD; Region: PHD-finger. PHD folds into an
                     interleaved type of Zn-finger chelating 2 Zn ions in a
                     similar manner to that of the RING and FYVE domains"
                     /db_xref="CDD:pfam00628"
ORIGIN

  Query Match             35.4%;  Score 1016.2;  DB 9;  Length 2767;
  Best Local Similarity   73.5%;  Pred. No. 7.2e-301;
  Matches 1414;  Conservative    0;  Mismatches  474;  Indels   36;  Gaps    8;

Qy       141 GCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTTTATGTAGGATTTCCAGCCAA 200
Db       192 GCTTGGACCCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTT-CAGGGAA 250

Qy       201 AGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGATAAAGGGCACCTGAGCAATGA 260
Db       251 AAGTGTCCAAGCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGA 310

Qy       261 GGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCG 320
Db       311 GGAGGCACCTAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCG 370

Qy       321 AAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACGGAGGCGGACCCCAGTGAGCA 380
Db       371 AAGGACCACTATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCC 430

Qy       381 GCAACC------ACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGACGGCAACCAAAACG 434
Db       431 GCCACCCCCACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCG 490

Qy       435 TACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGGAAAAAGAATGTGCC 494
Db       491 CACTGAGCGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCC 550

Qy       495 GGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACTGATGTGGAGACAGC 554
Db       551 TGTCTCCCTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGC 610

Qy       555 TTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCTGTATCTGACTCCTT 614
Db       611 CTCCGAGGGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCAC 670

Qy       615 A---GGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAAGAGGAAGAAGA 671
Db       671 AGCTGTGAAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGA 730

Qy       672 CACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAACCGCCTTCGGAG 731
Db       731 CACCTCCGATAGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAG 790

Qy       732 AAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAATCGCCTGAGGAA 791
Db       791 GAAGCGGGAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAA 850

Qy       792 GAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGCAGTGCCGAGCA 851
Db       851 GAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGA 910

Qy       852 GGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTGTCCCAGTCAGA 911
Db       911 GGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGG 970

Qy       912 GACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAATACAGAGGAAAA 971
Db       971 GAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGA 1030

Qy       972 CCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCCAATGCCCTGTA 1031
Db      1031 GCCTGGAGACTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTA 1090

Qy      1032 CTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGATCGGTGTGAGGA 1091
Db      1091 TTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGA 1150

Qy      1092 GTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAA 1151
Db      1151 ATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAA 1210

Qy      1152 CGGGGAAGACTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGATGAGACAAACGG 1211
Db      1211 TGGGGAAGACTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTC 1270

Qy      1212 TAGCGCCACCAATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGATGGCACAGACTG 1271
Db      1271 AGAAACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTG 1330

Qy      1272 CACAAGCATAGGGACAGTAGAGCAGAAGTCCGGAGAAGACCAGGGCATAAAGGGTAGGAT 1331
Db      1331 TACAAGTATAGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAAT 1390

Qy      1332 TGAGAAGGCAGCAAACCCCAGCGGCAAGAAAAAACTCAAGATATTCCAGCCTGTCGTAGA 1391
Db      1391 TGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGA 1450

Qy      1392 GGCTCCTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCACAGCCTGACTC 1451
Db      1451 GGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTC 1510

Qy      1452 TGTGTATTGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATGAGATTTCTAAG 1511
Db      1511 GGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAG 1570

Qy      1512 TTCAGGTAAAGAACAAAAAACAAAACCCAAGGAAAAGGTCAAGACGAAGCCAGAAAAGTT 1571
Db      1571 CTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCC 1630

Qy      1572 CAGTCTTCCAAAATGCAGTGTTCAGGTGGGGATTAAAATCTCTTCTGTGCACAAGAGACT 1631
```

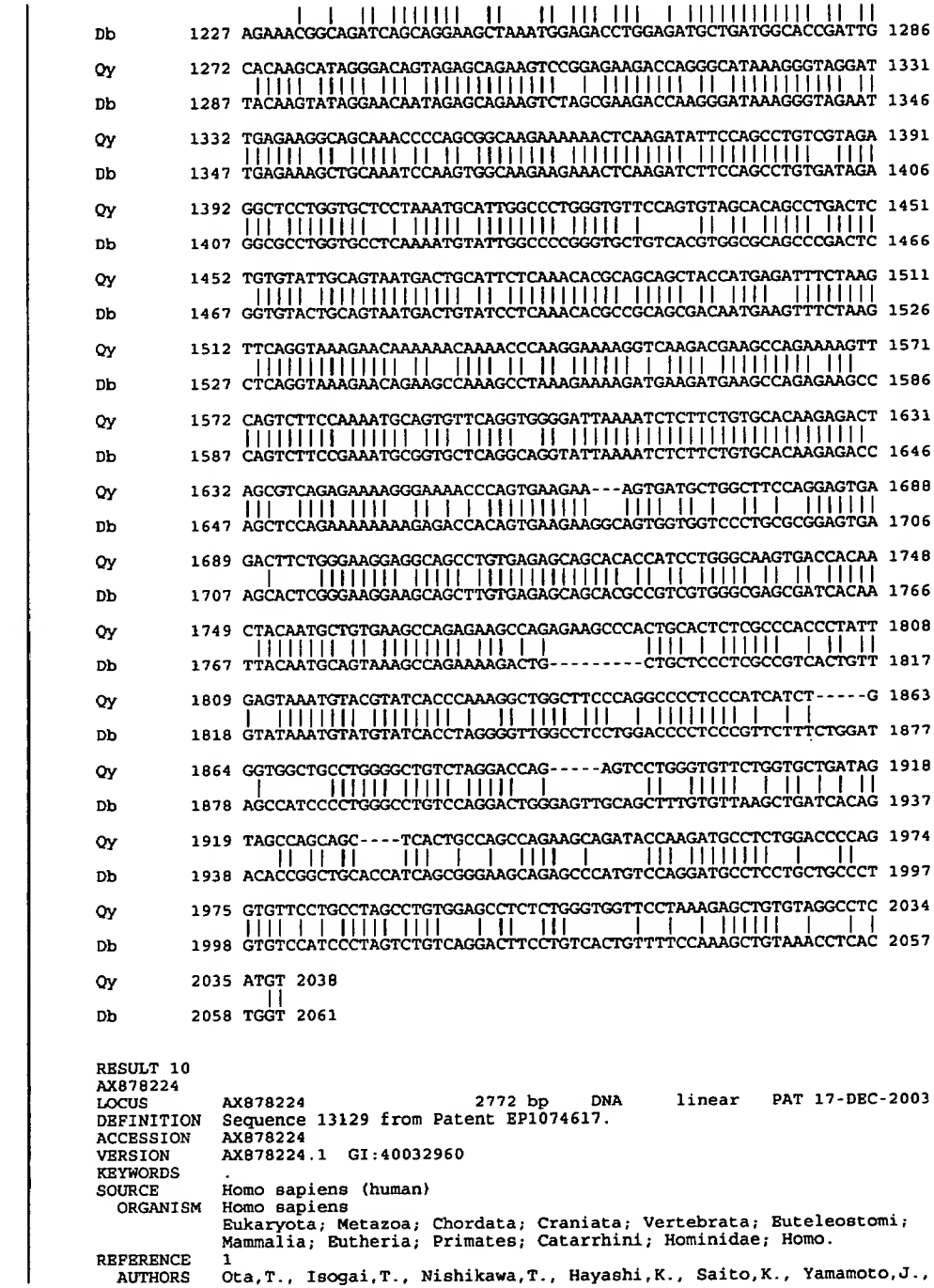
```
Db       1631 CAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACC 1690

Qy       1632 AGCGTCAGAGAAAAGGGAAAACCCAGTGAAGAA---AGTGATGCTGGCTTCCAGGAGTGA 1688

Db       1691 AGCTCCAGAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGA 1750

Qy       1689 GACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGCACACCATCCTGGGCAAGTGACCACAA 1748

Db       1751 AGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAA 1810

Qy       1749 CTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAGCCCACTGCACTCTCGCCCACCCTATT 1808

Db       1811 TTACAATGCAGTAAAGCCAGAAAAGACTG---------CTGCTCCCTCGCCGTCACTGTT 1861

Qy       1809 GAGTAAATGTACGTATCACCCAAAGGCTGGCTTCCCAGGCCCCTCCCATCATCT-----G 1863

Db       1862 GTATAAATGTATGTATCACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGAT 1921

Qy       1864 GGTGGCTGCCTGGGGCTGTCTAGGACCAG-----AGTCCTGGGTGTTCTGGTGCTGATAG 1918

Db       1922 AGCCATCCCCTGGGCCTGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAG 1981

Qy       1919 TAGCCAGCAGC----TCACTGCCAGCCAGAAGCAGATACCAAGATGCCTCTGGACCCCAG 1974

Db       1982 ACACCGGCTGCACCATCAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCT 2041

Qy       1975 GTGTTCCTGCCTAGCCTGTGGAGCCTCTCTGGGTGGTTCCTAAAGAGCTGTGTAGGCCTC 2034

Db       2042 GTGTCCATCCCTAGTCTGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCAC 2101

Qy       2035 ATGT 2038

Db       2102 TGGT 2105
```

RESULT 9
BC014489

```
LOCUS       BC014489                2725 bp    mRNA    linear   PRI 04-OCT-2003
DEFINITION  Homo sapiens death associated transcription factor 1, transcript
            variant 2, mRNA (cDNA clone MGC:23216 IMAGE:4901057), complete cds.
ACCESSION   BC014489
VERSION     BC014489.1  GI:15680266
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2725)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2725)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Pjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 34 Row: n Column: 1
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Similarity but not
            identity to protein.
FEATURES              Location/Qualifiers
     source           1. .2725
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="MGC:23216 IMAGE:4901057"
                      /tissue_type="Muscle, rhabdomyosarcoma"
                      /clone_lib="NIH_MGC_17"
                      /lab_host="DH10B-R"
                      /note="Vector: pOTB7"
     gene             1. .2725
                      /gene="DATF1"
                      /note="synonyms: DIO1, FLJ11265, KIAA0333"
                      /db_xref="LocusID:11083"
                      /db_xref="MIM:604140"
     CDS              238. .1926
                      /codon_start=1
                      /product="death associated transcription factor 1, isoform
                      a"
                      /protein_id="AAH14489.1"
                      /db_xref="GI:15680267"
                      /db_xref="LocusID:11083"
                      /translation="MDDKGDPSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDAE
                      ADPLEPPPPQQQLGLSLRRSGRQPKRTERVEQFLTIARRRGRRSMPVSLEDSGEPTSC
                      PATDAETASEGSVESASETRSGPQSASTAVKERPASSEKVKGGDDHDDTSDSDSDGLT
                      LKELQNRLRRKREQEPTERPLKGIQSRLRKKRREEGPAETVGSEASDTVEGVLPSKQE
                      PENDQGVVSQAGKDDRESKLEGKAAQDIKDEEPGDLGRPKPECEGYDPNALYCICRQP
                      HNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETHSETAD
                      QQEAKWRPGDADGTDCTSIGTIEQKSSEDQGIKGRIEKAANPSGKKKLKIFQPVIEAP
                      GASKCIGPGCCHVAQPDSVYCSNDCILKHAAATMKFLSSGKEQKPKPKEKMKMKPEKP
                      SLPKCGAQAGIKISSVHKRPAPEKKETTVKKAVVVPARSEALGKEAACESSTPSWASD
                      HNYNAVKPEKTAAPSPSLLYKCMYHLGVGLLDPSRSFWIAIPWACPGLGVAALC"
     misc_feature     1048. .1197
                      /note="PHD; Region: PHD-finger. PHD folds into an
                      interleaved type of Zn-finger chelating 2 Zn ions in a
                      similar manner to that of the RING and FYVE domains"
                      /db_xref="CDD:pfam00628"
ORIGIN
```

```
  Query Match           35.4%;  Score 1013.6;  DB 9;  Length 2725;
  Best Local Similarity  73.5%;  Pred. No. 4.5e-300;
  Matches 1414;  Conservative    0;  Mismatches  474;  Indels   36;  Gaps    8;
```





```
RESULT 10
AX878224
LOCUS       AX878224               2772 bp    DNA     linear   PAT 17-DEC-2003
DEFINITION  Sequence 13129 from Patent EP1074617.
ACCESSION   AX878224
VERSION     AX878224.1  GI:40032960
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
```

```
          Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
  TITLE   Primers for synthesising full-length cDNA and their use
  JOURNAL Patent: EP 1074617-A 13129 07-FEB-2001;
          Research Association for Biotechnology (JP)
FEATURES            Location/Qualifiers
     source         1. .2772
                    /organism="Homo sapiens"
                    /mol_type="unassigned DNA"
                    /db_xref="taxon:9606"
     CDS            301. .1989
                    /note="unnamed protein product"
                    /codon_start=1
                    /protein_id="CAE90061.1"
                    /db_xref="GI:40032961"
                    /translation="MDDKGDPSNEEAPKAIKPTSKEPRKTWGFRRTTIAKREGAGDAE
                    ADPLEPPPPQQQLGLSLRRSGRQPKRTERVEQFLTIARRRGRRSMPVSLEDSGEPTSC
                    PATDAETASEGSVESASETRSGPQSASTAVKERPASSEKVKGGDDHDDTSDSDSDGLT
                    LKELQNRLRRKREQEPTERPLKGIQSRLRKKRREEGPAETVGSEASDTVEGVLPSKQE
                    PENDQGVVSQAGKDDRESKLEGKAAQDIKDEEPEDLGRPKPECEGYDPNALYCICRQP
                    HNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETHSETAD
                    QQEAKWRPGDADGTDCTSIGTIEQKSSEDQGIKGRIEKAANPSGKKKLKIFQPVIEAP
                    GASKCIGPGCCHVAQPDSVYCSNDCILKHAAATMKFLSSGKEQKPKPKEKMKMKPEKP
                    SLPKCGAQAGIKISSVHKRPAPEKKETTVKKAVVVPARSEALGKEAACESSTPSWASD
                    HNYNAVKPEKTAAPSPSLLYKCMYHLGVGLLDPSRSFWIAIPWACPGLGVAALC"
ORIGIN

  Query Match           35.3%;  Score 1013;  DB 6;  Length 2772;
  Best Local Similarity 73.4%;  Pred. No. 7e-300;
  Matches 1413;  Conservative    0;  Mismatches  475;  Indels   36;  Gaps    8;

Qy       141 GCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTTTATGTAGGATTTCCAGCCAA 200
Db       211 GCTTGGACCCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTT-CAGGGAA 269

Qy       201 AGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGATAAAGGGCACCTGAGCAATGA 260
Db       270 AAGTGTCCAAGCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGA 329

Qy       261 GGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCG 320
Db       330 GGAGGCACCTAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCG 389

Qy       321 AAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACGGAGGCGGACCCCAGTGAGCA 380
Db       390 AAGGACCACTATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCC 449

Qy       381 GCAACC------ACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGACGGCAACCAAAACG 434
Db       450 GCCACCCCCACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCG 509

Qy       435 TACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGGAAAAAGAATGTGCC 494
Db       510 CACTGAGCGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCC 569

Qy       495 GGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACTGATGTGGAGACAGC 554
Db       570 TGTCTCCCTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGC 629

Qy       555 TTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCTGTATCTGACTCCTT 614
Db       630 CTCCGAGGGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCAC 689

Qy       615 A---GGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAAGAGGAAGAAGA 671
Db       690 AGCTGTGAAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGA 749

Qy       672 CACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAACCGCCTTCGGAG 731
Db       750 CACCTCCGATAGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAG 809

Qy       732 AAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAATCGCCTGAGGAA 791
Db       810 GAAGCGGGAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAA 869

Qy       792 GAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGCAGTGCCGAGCA 851
Db       870 GAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGA 929

Qy       852 GGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTGTCCCAGTCAGA 911
Db       930 GGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGG 989

Qy       912 GACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAATACAGAGGAAAA 971
Db       990 GAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGA 1049

Qy       972 CCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCCAATGCCCTGTA 1031
Db      1050 GCCTGAAGACTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTA 1109

Qy      1032 CTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGATCGGTGTGAGGA 1091
Db      1110 TTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGA 1169

Qy      1092 GTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAA 1151
Db      1170 ATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAA 1229

Qy      1152 CGGGGAAGACTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGATGAGACAAACGG 1211
Db      1230 TGGGGAAGACTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTC 1289

Qy      1212 TAGCGCCACCAATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGATGGCACAGACTG 1271
Db      1290 AGAAACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTG 1349

Qy      1272 CACAAGCATAGGGACAGTAGAGCAGAAGTCCGGAGAAGACCAGGGCATAAAGGGTAGGAT 1331
Db      1350 TACAAGTATAGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAAT 1409

Qy      1332 TGAGAAGGCAGCAAACCCCAGCGGCAAGAAAAAACTCAAGATATTCCAGCCTGTCGTAGA 1391
Db      1410 TGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGA 1469

Qy      1392 GGCTCCTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCACAGCCTGACTC 1451
Db      1470 GGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTC 1529

Qy      1452 TGTGTATTGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATGAGATTTCTAAG 1511
Db      1530 GGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAG 1589

Qy      1512 TTCAGGTAAAGAACAAAAAACAAAACCCAAGGAAAAGGTCAAGACGAAGCCAGAAAAGTT 1571
Db      1590 CTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCC 1649

Qy      1572 CAGTCTTCCAAAATGCAGTGTTCAGGTGGGGATTAAAATCTCTTCTGTGCACAAGAGACT 1631
Db      1650 CAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACC 1709

Qy      1632 AGCGTCAGAGAAAAGGGAAAACCCAGTGAAGAA---AGTGATGCTGGCTTCCAGGAGTGA 1688
Db      1710 AGCTCCAGAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGA 1769

Qy      1689 GACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGCACACCATCCTGGGCAAGTGACCACAA 1748
Db      1770 AGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAA 1829

Qy      1749 CTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAGCCCACTGCACTCTCGCCCACCCTATT 1808
Db      1830 TTACAATGCAGTAAAGCCAGAAAAGACTG---------CTGCTCCCTCGCCGTCACTGTT 1880

Qy      1809 GAGTAAATGTACGTATCACCCAAAGGCTGGCTTCCCAGGCCCCTCCCATCATCT-----G 1863
Db      1881 GTATAAATGTATGTATCACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGAT 1940
```

```
Qy       1864 GGTGGCTGCCTGGGGCTGTCTAGGACCAG-----AGTCCTGGGTGTTCTGGTGCTGATAG 1918
Db       1941 AGCCATCCCCTGGGCCTGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAG 2000

Qy       1919 TAGCCAGCAGC----TCACTGCCAGCCAGAAGCAGATACCAAGATGCCTCTGGACCCCAG 1974
Db       2001 ACACCGGCTGCACCATCAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCT 2060

Qy       1975 GTGTTCCTGCCTAGCCTGTGGAGCCTCTCTGGGTGGTTCCTAAAGAGCTGTGTAGGCCTC 2034
Db       2061 GTGTCCATCCCTAGTCTGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCAC 2120

Qy       2035 ATGT 2038
Db       2121 TGGT 2124
```

RESULT 11
BD157094

```
LOCUS       BD157094               2772 bp    DNA     linear   PAT 17-JAN-2003
DEFINITION  Primer for synthesizing full-length cDNA and use thereof.
ACCESSION   BD157094
VERSION     BD157094.1  GI:27862852
KEYWORDS    JP 2002191363-A/11937.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2772)
  AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
            Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
  TITLE     Primer for synthesizing full-length cDNA and use thereof
  JOURNAL   Patent: JP 2002191363-A 11937 09-JUL-2002;
            HELIX RESEARCH INSTITUTE
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002191363-A/11937
            PD   09-JUL-2002
            PF   28-JUL-2000 JP 2000280990
            PI   TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
            PI   SAITO,
            PI   JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
            PI   KEIICHI NAGAI,TETSUJI OTSUKI
            PC
            C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
             10,
            PC   C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
             Primer for synthesizing full-length cDNA and use thereof FH   Key
                      Location/Qualifiers
            FT   CDS             (301). .(1986).
FEATURES             Location/Qualifiers
     source          1. .2772
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN
```

```
  Query Match             35.3%;  Score 1013;  DB 6;  Length 2772;
  Best Local Similarity   73.4%;  Pred. No. 7e-300;
  Matches 1413;  Conservative    0;  Mismatches  475;  Indels   36;  Gaps    8;

Qy        141 GCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTTTATGTAGGATTTCCAGCCAA 200
Db        211 GCTTGGACCCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTT-CAGGGAA 269

Qy        201 AGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGATAAAGGGCACCTGAGCAATGA 260
Db        270 AAGTGTCCAAGCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGA 329

Qy        261 GGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCG 320
Db        330 GGAGGCACCTAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCG 389

Qy        321 AAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACGGAGGCGGACCCCAGTGAGCA 380
Db        390 AAGGACCACTATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCC 449

Qy        381 GCAACC------ACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGACGGCAACCAAAACG 434
Db        450 GCCACCCCCACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCG 509

Qy        435 TACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGGAAAAAGAATGTGCC 494
Db        510 CACTGAGCGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCC 569

Qy        495 GGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACTGATGTGGAGACAGC 554
Db        570 TGTCTCCCTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGC 629

Qy        555 TTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCTGTATCTGACTCCTT 614
Db        630 CTCCGAGGGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCAC 689

Qy        615 A---GGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAAGAGGAAGAAGA 671
Db        690 AGCTGTGAAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGA 749

Qy        672 CACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAACCGCCTTCGGAG 731
Db        750 CACCTCCGATAGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAG 809

Qy        732 AAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAATCGCCTGAGGAA 791
Db        810 GAAGCGGGAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAA 869

Qy        792 GAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGCAGTGCCGAGCA 851
Db        870 GAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGA 929

Qy        852 GGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTGTCCCAGTCAGA 911
Db        930 GGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGG 989

Qy        912 GACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAATACAGAGGAAAA 971
Db        990 GAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGA 1049

Qy        972 CCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCCAATGCCCTGTA 1031
Db       1050 GCCTGAAGACTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTA 1109

Qy       1032 CTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGATCGGTGTGAGGA 1091
Db       1110 TTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGA 1169

Qy       1092 GTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAA 1151
Db       1170 ATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAA 1229

Qy       1152 CGGGGAAGACTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGATGAGACAAACGG 1211
Db       1230 TGGGGAAGACTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTC 1289

Qy       1212 TAGCGCCACCAATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGATGGCACAGACTG 1271
Db       1290 AGAAACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTG 1349

Qy       1272 CACAAGCATAGGGACAGTAGAGCAGAAGTCCGGAGAAGACCAGGGCATAAAGGGTAGGAT 1331
Db       1350 TACAAGTATAGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAAT 1409

Qy       1332 TGAGAAGGCAGCAAACCCCAGCGGCAAGAAAAAACTCAAGATATTCCAGCCTGTCGTAGA 1391
Db       1410 TGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGA 1469

Qy       1392 GGCTCCTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCACAGCCTGACTC 1451
```

```
Db       1470 GGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTC 1529

Qy       1452 TGTGTATTGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATGAGATTTCTAAG 1511

Db       1530 GGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAG 1589

Qy       1512 TTCAGGTAAAGAACAAAAAACAAAACCCAAGGAAAAGGTCAAGACGAAGCCAGAAAAGTT 1571

Db       1590 CTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCC 1649

Qy       1572 CAGTCTTCCAAAATGCAGTGTTCAGGTGGGGATTAAAATCTCTTCTGTGCACAAGAGACT 1631

Db       1650 CAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACC 1709

Qy       1632 AGCGTCAGAGAAAAGGGAAAACCCAGTGAAGAA---AGTGATGCTGGCTTCCAGGAGTGA 1688

Db       1710 AGCTCCAGAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGA 1769

Qy       1689 GACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGCACACCATCCTGGGCAAGTGACCACAA 1748

Db       1770 AGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAA 1829

Qy       1749 CTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAGCCCACTGCACTCTCGCCCACCCTATT 1808

Db       1830 TTACAATGCAGTAAAGCCAGAAAAGACTG---------CTGCTCCCTCGCCGTCACTGTT 1880

Qy       1809 GAGTAAATGTACGTATCACCCAAAGGCTGGCTTCCCAGGCCCCTCCCATCATCT-----G 1863

Db       1881 GTATAAATGTATGTATCACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGAT 1940

Qy       1864 GGTGGCTGCCTGGGGCTGTCTAGGACCAG-----AGTCCTGGGTGTTCTGGTGCTGATAG 1918

Db       1941 AGCCATCCCCTGGGCCTGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAG 2000

Qy       1919 TAGCCAGCAGC----TCACTGCCAGCCAGAAGCAGATACCAAGATGCCTCTGGACCCCAG 1974

Db       2001 ACACCGGCTGCACCATCAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCT 2060

Qy       1975 GTGTTCCTGCCTAGCCTGTGGAGCCTCTCTGGGTGGTTCCTAAAGAGCTGTGTAGGCCTC 2034

Db       2061 GTGTCCATCCCTAGTCTGTCAGGACTTCCTGTCACTGTTTCCAAAGCTGTAAACCTCAC 2120

Qy       2035 ATGT 2038

Db       2121 TGGT 2124
```

RESULT 12
AK002127

```
LOCUS       AK002127               2772 bp    mRNA    linear   PRI 01-AUG-2002
DEFINITION  Homo sapiens cDNA FLJ11265 fis, clone PLACE1009158.
ACCESSION   AK002127
VERSION     AK002127.1  GI:7023814
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
            Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
            Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
            Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
            Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
            Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2772)
  AUTHORS   Isogai,T. and Otsuki,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
            Helix Research Institute (supported by Japan Key Technology Center
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES             Location/Qualifiers
     source          1. .2772
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="PLACE1009158"
                     /tissue_type="placenta"
                     /clone_lib="PLACE1"
                     /note="cloning vector: pME18SFL3"
     CDS             301. .1989
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAA92094.1"
                     /db_xref="GI:7023815"
                     /translation="MDDKGDPSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDAE
                     ADPLEPPPPQQQLGLSLRRSGRQPKRTERVEQFLTIARRRGRRSMPVSLEDSGEPTSC
                     PATDAETASEGSVESASETRSGPQSASTAVKERPASSEKVKGGDDHDDTSDSDSDGLT
                     LKELQNRLRRKREQEPTERPLKGIQSRLRKKRREEGPAETVGSEASDTVEGVLPSKQE
                     PENDQGVVSQAGKDDRESKLEGKAAQDIKDEEPEDLGRPKPECEGYDPNALYCICRQP
                     HNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETHSETAD
                     QQEAKWRPGDADGTDCTSIGTIEQKSSEDQGIKGRIEKAANPSGKKKLKIFQPVIEAP
                     GASKCIGPGCCHVAQPDSVYCSNDCILKHAAATMKFLSSGKEQKPKPKEKMKMKPEKP
                     SLPKCGAQAGIKISSVHKRPAPEKKETTVKKAVVVPARSEALGKEAACESSTPSWASD
                     HNYNAVKPEKTAAPSPSLLYKCMYHLGVGLLDPSRSFWIAIPWACPGLGVAALC"
ORIGIN
```

```
  Query Match           35.3%;  Score 1013;  DB 9;  Length 2772;
  Best Local Similarity  73.4%;  Pred. No. 7e-300;
  Matches 1413;  Conservative    0;  Mismatches  475;  Indels   36;  Gaps    8;

Qy        141 GCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTTTATGTAGGATTTCCAGCCAA 200

Db        211 GCTTGGACCCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTT-CAGGGAA 269

Qy        201 AGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGATAAAGGGCACCTGAGCAATGA 260

Db        270 AAGTGTCCAAGCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGA 329

Qy        261 GGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCG 320

Db        330 GGAGGCACCTAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCG 389

Qy        321 AAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACGGAGGCGGACCCCAGTGAGCA 380

Db        390 AAGGACCACTATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCC 449

Qy        381 GCAACC------ACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGACGGCAACCAAAACG 434

Db        450 GCCACCCCCACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGGAGGCAGCCCAAGCG 509

Qy        435 TACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGGAAAAAGAATGTGCC 494

Db        510 CACTGAGCGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCC 569

Qy        495 GGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACTGATGTGGAGACAGC 554

Db        570 TGTCTCCCTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGC 629

Qy        555 TTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCTGTATCTGACTCCTT 614

Db        630 CTCCGAGGGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCAC 689

Qy        615 A---GGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAAGAGGAAGAAGA 671

Db        690 AGCTGTGAAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGA 749
```

```
Qy       672 CACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAACCGCCTTCGGAG 731
Db       750 CACCTCCGATAGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAG 809

Qy       732 AAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAATCGCCTGAGGAA 791
Db       810 GAAGCGGGAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAA 869

Qy       792 GAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGCAGTGCCGAGCA 851
Db       870 GAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGA 929

Qy       852 GGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTGTCCCAGTCAGA 911
Db       930 GGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGG 989

Qy       912 GACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAATACAGAGGAAAA 971
Db       990 GAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGA 1049

Qy       972 CCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCCAATGCCCTGTA 1031
Db      1050 GCCTGAAGACTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTA 1109

Qy      1032 CTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGATCGGTGTGAGGA 1091
Db      1110 TTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGA 1169

Qy      1092 GTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAA 1151
Db      1170 ATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAA 1229

Qy      1152 CGGGGAAGACTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGATGAGACAAACGG 1211
Db      1230 TGGGGAAGACTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTC 1289

Qy      1212 TAGCGCCACCAATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGATGGCACAGACTG 1271
Db      1290 AGAAACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTG 1349

Qy      1272 CACAAGCATAGGGACAGTAGAGCAGAAGTCCGGAGAAGACCAGGGCATAAAGGGTAGGAT 1331
Db      1350 TACAAGTATAGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAAT 1409

Qy      1332 TGAGAAGGCAGCAAACCCCAGCGGCAAGAAAAAACTCAAGATATTCCAGCCTGTCGTAGA 1391
Db      1410 TGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGA 1469

Qy      1392 GGCTCCTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCACAGCCTGACTC 1451
Db      1470 GGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTC 1529

Qy      1452 TGTGTATTGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATGAGATTTCTAAG 1511
Db      1530 GGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAG 1589

Qy      1512 TTCAGGTAAAGAACAAAAAACAAAACCCAAGGAAAAGGTCAAGACGAAGCCAGAAAAGTT 1571
Db      1590 CTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCC 1649

Qy      1572 CAGTCTTCCAAAATGCAGTGTTCAGGTGGGGATTAAAATCTCTTCTGTGCACAAGAGACT 1631
Db      1650 CAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACC 1709

Qy      1632 AGCGTCAGAGAAAAGGGAAAACCCAGTGAAGAA---AGTGATGCTGGCTTCCAGGAGTGA 1688
Db      1710 AGCTCCAGAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGA 1769

Qy      1689 GACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGCACACCATCCTGGGCAAGTGACCACAA 1748
Db      1770 AGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAA 1829

Qy      1749 CTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAGCCCACTGCACTCTCGCCCACCCTATT 1808
Db      1830 TTACAATGCAGTAAAGCCAGAAAAGACTG---------CTGCTCCCTCGCCGTCACTGTT 1880

Qy      1809 GAGTAAATGTACGTATCACCCAAAGGCTGGCTTCCCAGGCCCCTCCCATCATCT-----G 1863
Db      1881 GTATAAATGTATGTATCACCTAGGGGTTGGCCTCCTGGACCCCTCCCGTTCTTTCTGGAT 1940

Qy      1864 GGTGGCTGCCTGGGGCTGTCTAGGACCAG-----AGTCCTGGGTGTTCTGGTGCTGATAG 1918
Db      1941 AGCCATCCCCTGGGCCTGTCCAGGACTGGGAGTTGCAGCTTTGTGTTAAGCTGATCACAG 2000

Qy      1919 TAGCCAGCAGC----TCACTGCCAGCCAGAAGCAGATACCAAGATGCCTCTGGACCCCAG 1974
Db      2001 ACACCGGCTGCACCATCAGCGGGAAGCAGAGCCCATGTCCAGGATGCCTCCTGCTGCCCT 2060

Qy      1975 GTGTTCCTGCCTAGCCTGTGGAGCCTCTCTGGGTGGTTCCTAAAGAGCTGTGTAGGCCTC 2034
Db      2061 GTGTCCATCCCTAGTCTGTCAGGACTTCCTGTCACTGTTTTCCAAAGCTGTAAACCTCAC 2120

Qy      2035 ATGT 2038
Db      2121 TGGT 2124
```

```
RESULT 13
BC000770
LOCUS       BC000770               2407 bp    mRNA    linear   PRI 12-NOV-2003
DEFINITION  Homo sapiens death associated transcription factor 1, transcript
            variant 3, mRNA (cDNA clone MGC:3257 IMAGE:3506207), complete cds.
ACCESSION   BC000770
VERSION     BC000770.2  GI:33990873
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2407)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2407)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2000) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     On Aug 20, 2003 this sequence version replaced gi:12653952.
            Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
```

```
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 7 Row: i Column: 2
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, Similarity but not identity to protein.
FEATURES             Location/Qualifiers
     source          1..2407
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:3257 IMAGE:3506207"
                     /tissue_type="Kidney, renal cell adenocarcinoma"
                     /clone_lib="NIH_MGC_14"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     gene            1..2407
                     /gene="DATF1"
                     /note="synonyms: DIO1, FLJ11265, KIAA0333"
                     /db_xref="LocusID:11083"
                     /db_xref="MIM:604140"
     CDS             201..1835
                     /codon_start=1
                     /product="death associated transcription factor 1, isoform
                     b"
                     /protein_id="AAH00770.1"
                     /db_xref="GI:12653953"
                     /db_xref="LocusID:11083"
                     /translation="MDDKGDPSNEEAPKAIKPTSKEFRKTWGFRRTTIAKREGAGDAE
                     ADPLEPPPPQQQLGLSLRRSGRQPKRTERVEQFLTIARRRGRRSMPVSLEDSGEPTSC
                     PATDAETASEGSVESASETRSGPQSASTAVKERPASSEKVKGGDDHDDTSDSDSDGLT
                     LKELQNRLRRKREQEPTERPLKGIQSRLRKKRREEGPAETVGSEASDTVEGVLPSKQE
                     PENDQGVVSQAGKDDRESKLEGKAAQDIKDEEPGDLGRPKPECEGYDPNALYCICRQP
                     HNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTILQVQDETHSETAD
                     QQEAKWRPGDADGTDCTSIGTIEQKSSEDQGIKGRIEKAANPSGKKKLKIFQPVIEAP
                     GASKCIGPGCCHVAQPDSVYCSNDCILKHAAATMKFLSSGKEQKPKPKEKMKMKPEKP
                     SLPKCGAQAGIKISSVHKRPAPEKKETTVKKAVVVPARSEALGKEAACESSTPSWASD
                     HNYNAVKPEKTAAPSPSLLYKCSGKYLYSLHPSLIA"
     misc_feature    1011..1160
                     /note="PHD; Region: PHD-finger. PHD folds into an
                     interleaved type of Zn-finger chelating 2 Zn ions in a
                     similar manner to that of the RING and FYVE domains"
                     /db_xref="CDD:pfam00628"
ORIGIN

  Query Match             34.4%;  Score 985;  DB 9;  Length 2407;
  Best Local Similarity   76.3%;  Pred. No. 2.9e-291;
  Matches 1266;  Conservative    0;  Mismatches  380;  Indels   13;  Gaps    4;

Qy        141 GCTTGGCCGTCTGCTTGACCTGGCCCCAGGTCTATAATTTTATGTAGGATTTCCAGCCAA 200
Db        111 GCTTGGACCCAGAGGTTTCGTCTGCCAGGGTTTTTGGTTGTATTTAGGATTT-CAGGGAA 169

Qy        201 AGGTTTCCAAGCTTTCAGTTTTGGGACAGGTATGGATGATAAAGGGCACCTGAGCAATGA 260
Db        170 AAGTGTCCAAGCTTTCAGTGTTGGAGCAGGTATGGACGACAAAGGCGACCCGAGCAATGA 229

Qy        261 GGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCG 320
Db        230 GGAGGCACCTAAGGCCATCAAACCCACCAGCAAAGAGTTCAGGAAAACATGGGGTTTTCG 289

Qy        321 AAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACGGAGGCGGACCCCAGTGAGCA 380
Db        290 AAGGACCACTATCGCCAAGCGAGAGGGCGCAGGGGACGCGGAGGCTGACCCACTGGAGCC 349

Qy        381 GCAACC------ACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGACGGCAACCAAAACG 434
Db        350 GCCACCCCCACAGCAGCAGCTGGGCCTGTCCCTGCGGCGCAGTGGAGGCAGCCCAAGCG 409

Qy        435 TACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGGAAAAAGAATGTGCC 494
Db        410 CACTGAGCGCGTGGAGCAGTTCCTGACCATTGCGCGGCGCCGCGGCAGGAGGAGCATGCC 469

Qy        495 GGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACTGATGTGGAGACAGC 554
Db        470 TGTCTCCCTGGAGGATTCTGGTGAGCCCACGTCCTGCCCCGCCACAGACGCCGAGACAGC 529

Qy        555 TTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCTGTATCTGACTCCTT 614
Db        530 CTCCGAGGGCAGCGTGGAAAGCGCTTCTGAGACCAGAAGCGGCCCCCAGTCTGCTTCCAC 589

Qy        615 AGG---GAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAAGAGGAAGAAGA 671
Db        590 AGCTGTGAAGGAACGACCAGCCTCTTCTGAAAAGGTGAAAGGAGGGGATGACCACGATGA 649

Qy        672 CACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAACCGCCTTCGGAG 731
Db        650 CACCTCCGATAGTGACAGCGATGGCCTGACCTTGAAAGAGCTTCAGAATCGCCTTCGCAG 709

Qy        732 AAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAATCGCCTGAGGAA 791
Db        710 GAAGCGGGAACAGGAGCCCACTGAGAGGCCCCTGAAAGGGATCCAGAGTCGCCTGCGGAA 769

Qy        792 GAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGCAGTGCCGAGCA 851
Db        770 GAAGCGCCGGGAGGAGGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGA 829

Qy        852 GGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTGTCCCAGTCAGA 911
Db        830 GGGCGTCCTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCCAGGCTGG 889

Qy        912 GACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAATACAGAGGAAAA 971
Db        890 GAAAGATGACAGAGAGAGTAAGTTGGAGGGAAAGGCGGCTCAGGACATCAAAGATGAGGA 949

Qy        972 CCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCCAATGCCCTGTA 1031
Db        950 GCCTGGAGACTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACGCCCTGTA 1009

Qy       1032 CTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGATCGGTGTGAGGA 1091
Db       1010 TTGCATTTGCCGCCAGCCTCACAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGA 1069

Qy       1092 GTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGCTCCTGGAAAGGAA 1151
Db       1070 ATGGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAA 1129

Qy       1152 CGGGGAAGACTACATCTGCCCAAATTGCACCATTTTGCAAGTGCAGGATGAGACAAACGG 1211
Db       1130 TGGGGAAGACTATATCTGCCCAAACTGCACCATTCTGCAAGTGCAGGATGAGACTCATTC 1189

Qy       1212 TAGCGCCACCAATGAGCAGGACTCTGGGTGCAGATCTGTGGGTGCTGATGGCACAGACTG 1271
Db       1190 AGAAACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTG 1249

Qy       1272 CACAAGCATAGGGACAGTAGAGCAGAAGTCCGGAGAAGACCAGGGCATAAAGGGTAGGAT 1331
Db       1250 TACAAGTATAGGAACAATAGAGCAGAAGTCTAGCGAAGACCAAGGGATAAAGGGTAGAAT 1309

Qy       1332 TGAGAAGGCAGCAAACCCCAGCGGCAAGAAAAAACTCAAGATATTCCAGCCTGTCGTAGA 1391
Db       1310 TGAGAAAGCTGCAAATCCAAGTGGCAAGAAGAAACTCAAGATCTTCCAGCCTGTGATAGA 1369

Qy       1392 GGCTCCTGGTGCTCCTAAATGCATTGGCCCTGGGTGTTCCAGTGTAGCACAGCCTGACTC 1451
Db       1370 GGCGCCTGGTGCCTCAAAATGTATTGGCCCCGGGTGCTGTCACGTGGCGCAGCCCGACTC 1429

Qy       1452 TGTGTATTGCAGTAATGACTGCATTCTCAAACACGCAGCAGCTACCATGAGATTTCTAAG 1511
```

```
Db       1430 GGTGTACTGCAGTAATGACTGTATCCTCAAACACGCCGCAGCGACAATGAAGTTTCTAAG 1489

Qy       1512 TTCAGGTAAAGAACAAAAAACAAAACCCAAGGAAAAGGTCAAGACGAAGCCAGAAAAGTT 1571

Db       1490 CTCAGGTAAAGAACAGAAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAAGCC 1549

Qy       1572 CAGTCTTCCAAAATGCAGTGTTCAGGTGGGGATTAAAATCTCTTCTGTGCACAAGAGACT 1631

Db       1550 CAGTCTTCCGAAATGCGGTGCTCAGGCAGGTATTAAAATCTCTTCTGTGCACAAGAGACC 1609

Qy       1632 AGCGTCAGAGAAAAGGGAAAACCCAGTGAAGAA---AGTGATGCTGGCTTCCAGGAGTGA 1688

Db       1610 AGCTCCAGAAAAAAAAGAGACCACAGTGAAGAAGGCAGTGGTGGTCCCTGCGCGGAGTGA 1669

Qy       1689 GACTTCTGGGAAGGAGGCAGCCTGTGAGAGCAGCACACCATCCTGGGCAAGTGACCACAA 1748

Db       1670 AGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACGCCGTCGTGGGCGAGCGATCACAA 1729

Qy       1749 CTACAATGCTGTGAAGCCAGAGAAGCCAGAGAAGCCCAC 1787

Db       1730 TTACAATGCAGTAAAGCCAGAAAAGACTGCTGCTCCCTC 1768
```

```
RESULT 14
AC119716
LOCUS       AC119716               221547 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-274I16, *** SEQUENCING IN PROGRESS
            ***, 17 unordered pieces.
ACCESSION   AC119716
VERSION     AC119716.6  GI:25092366
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 221547)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 221547)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 221547)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 19, 2002 this sequence version replaced gi:23611286.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GWLC
                Center clone name: CH230-274I16
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 208013 bases at least Q40
                Consensus quality: 211685 bases at least Q30
                Consensus quality: 213907 bases at least Q20
                Estimated insert size: 199182; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   198201: contig of 198201 bp in length
            *  198202   198301: gap of unknown length
            *  198302   199749: contig of 1448 bp in length
            *  199750   199849: gap of unknown length
            *  199850   201115: contig of 1266 bp in length
            *  201116   201215: gap of unknown length
```

```
        *    201216    202338: contig of 1123 bp in length
        *    202339    202438: gap of unknown length
        *    202439    203500: contig of 1062 bp in length
        *    203501    203600: gap of unknown length
        *    203601    204905: contig of 1305 bp in length
        *    204906    205005: gap of unknown length
        *    205006    206192: contig of 1187 bp in length
        *    206193    206292: gap of unknown length
        *    206293    207385: contig of 1093 bp in length
        *    207386    207485: gap of unknown length
        *    207486    208519: contig of 1034 bp in length
        *    208520    208619: gap of unknown length
        *    208620    209758: contig of 1139 bp in length
        *    209759    209858: gap of unknown length
        *    209859    211441: contig of 1583 bp in length
        *    211442    211541: gap of unknown length
        *    211542    212876: contig of 1335 bp in length
        *    212877    212976: gap of unknown length
        *    212977    214256: contig of 1280 bp in length
        *    214257    214356: gap of unknown length
        *    214357    217003: contig of 2647 bp in length
        *    217004    217103: gap of unknown length
        *    217104    218459: contig of 1356 bp in length
        *    218460    218559: gap of unknown length
        *    218560    219788: contig of 1229 bp in length
        *    219789    219888: gap of unknown length
        *    219889    221547: contig of 1659 bp in length.
FEATURES             Location/Qualifiers
     source          1..221547
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-274I16"
     misc_feature    1..1272
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    complement(4416..5263)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end_sequence:BZ266352"
     misc_feature    192445..193553
                     /note="wgs_contig"
     misc_feature    195743..196592
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BZ266349"
ORIGIN

  Query Match             29.8%;  Score 853.4;  DB 2;  Length 221547;
  Best Local Similarity   81.7%;  Pred. No. 2.1e-250;
  Matches  979;  Conservative    0;  Mismatches  107;  Indels  112;  Gaps    1;

Qy         53 TTAAGAGGCAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGAT 112
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     206293 TTAAGAGGCAGTCCCCACTGTCCCTTTGGTTGTTCGAAGCTCCGGAATCTTCTCATTGAT 206352

Qy        113 GAACTGTGACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTC 172
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     206353 GAACTGTGACTGACAACAGTGGGGTGAGGCTTGGCCGTCTGCTTGACCTGGCCCCAGGTC 206412

Qy        173 TATAATTTTATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGAC----- 227
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     206413 TATAATTTTATGTAGGATTTCCAGCCAAAGGTTTCCAAGCTTTCAGTTTTGGGACAGGGT 206472

Qy        228 ------------------------------------------------------------ 227

Db     206473 TTTTGTTTTTATGCCAGTCTCCCCCTAAGATGGAACATGCAGTAAACTTGTCCATTACTA 206532

Qy        228 -----------------------------------------------AGGTATGGATGAT 240
                                                             |||||||||||||
Db     206533 GCATTTAAATGCAGTTAAAGCCTGGACAGCCAGGAAAAGAATTCTCAAGGTATGGATGAT 206592

Qy        241 AAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     206593 AAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCACCAGTAAGGAGTTC 206652

Qy        301 AGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACG 360
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db     206653 AGGAAAACCTGNGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGGTGCAGGAGACACG 206712

Qy        361 GAGGCGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGA 420
              |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     206713 GAGGTGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCTCCCTGCGCCGCAGTGGA 206772

Qy        421 CGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGTTCGGCGCCGAGGG 480
              ||||||||||||||||||||||||||||||||||||   |||||||||||||||||||||
Db     206773 CGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTTACACGGTTCGGCGCCGAGGG 206832

Qy        481 AAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     206833 AAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATCTTCCACAGTCACT 206892

Qy        541 GATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     206893 GATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGATCAGAAGTGGCCCT 206952

Qy        601 GTATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     206953 GTATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAAAAGGCAAAAGGAGGTGAA 207012

Qy        661 GAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAAC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     207013 GAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACGTTGAAGGAACTTCAGAAC 207072

Qy        721 CGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     207073 CGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAGAGGCAGTCAGAAT 207132

Qy        781 CGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     207133 CGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAGTGTCCAAATAGGC 207192

Qy        841 AGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     207193 AGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAGTCAGGGACCAGTG 207252

Qy        901 TCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     207253 TCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGAAGGCGACTCAGGGAAAT 207312

Qy        961 ACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     207313 ACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGAGGTTTACGACCCC 207372

Qy       1021 AATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGTTTATGATCTGCTGTGAT 1080
              ||||||||||||
Db     207373 AATGCCCTGTACTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 207432

Qy       1081 CGGTGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATTTCTGAGGCCCGAGGGCGGC 1138
                                                                     | |
Db     207433 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCAGCC 207490


RESULT 15
AC108337
LOCUS       AC108337               272404 bp    DNA     linear   HTG 21-SEP-2002
DEFINITION  Rattus norvegicus clone CH230-115I13, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION   AC108337
VERSION     AC108337.4  GI:23266147
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 272404)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,P., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,P., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 272404)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (27-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 272404)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (21-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Sep 21, 2002 this sequence version replaced gi:21737648.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequening reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GPMX
Center clone name: CH230-115I13.
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 238513 bases at least Q40
Consensus quality: 241910 bases at least Q30
Consensus quality: 244206 bases at least Q20
Estimated insert size: 262154; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 270622: contig of 270622 bp in length
* 270623 270722: gap of unknown length
* 270723 272404: contig of 1682 bp in length.

FEATURES Location/Qualifiers
source 1. .272404
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-115I13"
misc_feature 1. .1256
/note="wgs_end_extension
clone_end:Sp6"
misc_feature complement(1814. .2663)
/note="clone_boundary
clone_end:Sp6
site:EcoRI
end_sequence:BH294125"
misc_feature 62809. .64046
/note="wgs_contig"
misc_feature 267959. .268006
/note="clone_boundary
clone_end:T7
site:EcoRI
end_sequence:BH294091"

ORIGIN

Query Match 26.1%; Score 748.2; DB 2; Length 272404;
Best Local Similarity 92.1%; Pred. No. 5.8e-218;
Matches 789; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```
Qy        228 AGGTATGGATGATAAAGGGCACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCAC 287
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     128279 AGGTATGGATGATAAAGGGGACCTGAGCAATGAGGAAGCACCCAAGGCTATCAAACCCAC 128338

Qy        288 CAGTAAGGAGTTCAGGAAAACCTGGGGTTTTCGAAGAACCACGATTGCCAAACGTGAGGG 347
              |||||| |||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     128339 CAGTAAAGAGTTCAGGAAAACCTGGGGTTTTCGAAGAACCACAATTGCCAAACGTGAGGG 128398

Qy        348 TGCAGGAGACACGGAGGCGGACCCCAGTGAGCAGCAACCACAGCAGCATAACCTCTCCCT 407
              |||||||||||| |||| |||||||||||||||||||||||| ||||||||||||||||||
Db     128399 TGCAGGAGACACCGAGGTGGACCCCAGTGAGCAGCAACCACAACAGCATAACCTCTCCCT 128458

Qy        408 GCGCCGCAGTGGACGGCAACCAAAACGTACTGAGAGGGTAGAAGAGTTTCTTACCACGGT 467
              |||||||||||||||||| ||||| || || || |||||||||||||| ||||| || ||
Db     128459 GCGCCGCAGTGGACGGCAGCCAAAGCGGACCGAAAGGGTAGAAGAGTTCCTTACTACTGT 128518
```

```
Qy       468 TCGGCGCCGAGGGAAAAAGAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACATC 527
Db    128519 TCGGCGCCGAGGGAGAAAAAATGTGCCGGTGTCCCTGGAGGATTCCAGTGAGCCCACGTC 128578

Qy       528 TTCCACAGTCACTGATGTGGAGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAGAT 587
Db    128579 CTCCACAGTCACTGATGTGGAGACAGCTTCTGAGGGGAGCGTTGAAAGCAGTTCTGAGAT 128638

Qy       588 CAGAAGTGGCCCTGTATCTGACTCCTTAGGGAAAGAACATCCTGCCTCTTCTGAAAAGGC 647
Db    128639 CAGAAGTGGCCCTGCACCTGGCACCTTAGGGAAAGAACATCCTTCCTCTTCTGAAAAGGT 128698

Qy       648 AAAAGGAGGTGAAGAGGAAGAAGACACCTCTGACAGTGACAGTGATGGCCTTACGTTGAA 707
Db    128699 GAAAGGAGGTGATGAGGAAGAAGACACCTCTGACAGTGACAGTGATGGTCTTACGCTGAA 128758

Qy       708 GGAACTTCAGAACCGCCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCCCTGAG 767
Db    128759 GGAACTTCAGAACCGCCTTCGAAGAAAGCGAGAGCAGGAACCTGTGGAGAGGTCCCTGAG 128818

Qy       768 AGGCAGTCAGAATCGCCTGAGGAAGAAGCGCAGAGAGGAAGATTCTGCCGAAACTGGGAG 827
Db    128819 AGGCATTCAGAATCGACTGCGGAAGCAGCGCCGAGAGGAAGATCCTGCCGAAGCTGGGAG 128878

Qy       828 TGTCCAAATAGGCAGTGCCGAGCAGGACAGACCTCTCTGTAAGCAGGAGCCTGAGGCTAG 887
Db    128879 TGTCCAGATAGGCAGTATTGAGGAGGACACACCTCGCTGTAAGCAGGAGCCTGAGGCTAG 128938

Qy       888 TCAGGGACCAGTGTCCCAGTCAGAGACAGATGACATAGAAAATCAGTTGGAAGGGAAGGC 947
Db    128939 CCAGGGACCAGTGTCCCAGTCAGAGAAAGATGACAGAGAAAATCAGTTGGAGGGGAAAGC 128998

Qy       948 GACTCAGGGAAATACAGAGGAAAACCCCAGGGAAGCGGGCAAACCAAAGCCTGAGTGTGA 1007
Db    128999 GACTCAGGGAAGTAAGGAGGAAAGCCCCAGGGAAGCTGGCAAACCAAAGCCTGAGTGTGA 129058

Qy      1008 GGTTTACGACCCCAATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGTTTAT 1067
Db    129059 GGTTTACGACCCCAATGCCCTGTACTGCATCTGCCGCCAGCCTCACAACAACAGGTAGGT 129118

Qy      1068 GATCTGCTGTGATCGGT 1084
Db    129119 GGAGGGAGGGAACTGGT 129135
```

Search completed: April 23, 2004, 14:16:35
Job time : 10938.8 secs